Journal of Research in Reading, ISSN 0141-0423 *DOI: 10.1111/j.1467-9817.2008.01383.x*
Volume 32, Issue 1, 2009, pp 92–108

A randomised efficacy study of Web-based synthetic and analytic programmes among disadvantaged urban Kindergarten children

Erin M. Comaskey and Robert S. Savage
McGill University, Canada

Philip Abrami
Centre for the Study of Learning and Performance, Concordia University, Canada

This study explores whether two computer-based literacy interventions – a 'synthetic phonics' and an 'analytic phonics' approach produce qualitatively distinct effects on the early phonological abilities and reading skills of disadvantaged urban Kindergarten (Reception) children. Participants ($n = 53$) were assigned by random allocation to one of the two interventions. Each intervention was generally delivered three times per week for 13 weeks as part of a reading centre approach in Kindergarten classrooms with small groups of children. In the synthetic programme children showed, as predicted, significant ($p < .05$) improvement in CV and VC word blending and the articulation of final consonants. The children in the analytic phonics programme showed, as predicted, significant ($p < .05$) improvements in articulating shared rimes in words. These results suggest that synthetic and analytic programmes have qualitatively different effects on children's phonological development. These phonological differences are not however immediately reflected in any qualitative differences in the way children undertook word reading or nonword decoding.

The purposes of this paper are twofold. Firstly, we briefly review the literature on the effectiveness of computer-based reading interventions and all intervention studies exploring the efficacy of synthetic and analytic phonics programmes. Secondly, we identify the need for, and then report the results of, a randomised intervention study using computers to deliver Web-based phonics interventions. We explore whether qualitatively different effects of synthetic and analytic phonics programmes were evident for disadvantaged children in an urban school.

The evidence base for computer-based literacy interventions

What does the best evidence accumulated to date say on the effectiveness of computer-based literacy interventions? The most comprehensive review of studies of the impact of information and communication technology (ICT) on literacy was conducted by Torgerson and Zhu (2003). This meta-analysis used explicit criteria to select studies, limiting inclusion to randomised control trials (RCTs) of children aged 5–16. Twelve studies included for in-depth review were pared down from over 2,000 candidate studies initially identified on ERIC and PsychInfo databases. Analysis of these 12 studies showed mixed effects of interventions. Overall, effect sizes were not significantly different from zero. Torgerson and Zhu concluded firstly that teachers and policy-makers should be aware that there is not sufficient high-quality evidence available to justify the use of ICT to support literacy, and secondly that rigorously designed RCT studies of the impact of technology on literacy are urgently required. Similar results have been reported from a recent large-scale study in the United States by Dynarski et al. (2007). An RCT study of a range of commercially available ICT products was undertaken in 132 schools in 33 school districts and with a total of 4,389 teachers. Dynarski et al. showed that the effect size for interventions was not significantly different from zero on standard and local tests of reading in either Grade 1 or Grade 4.

Few studies to date however have explored the effects of technology in Kindergarten. It is also not clear to what extent the effects reported above reflect weaknesses in implementation of the individual programmes themselves rather than a more fundamental problem of ICT effectiveness. For example, in many studies teachers are given a single day's training in using technology and many have reported being unsure about using it in the classroom (e.g. Dynarski et al., 2007). Interventions run by well-trained facilitators rather than briefly trained teachers to establish *efficacy* of ICTs might be an important first step.

As Torgerson and Zhu identify, there is clearly a need for more randomised studies in this domain. In the medical sciences, there is growing consensus not only that well-designed RCT studies are needed but that such studies should be executed and reported following clear and consistent criteria such as the CONSORT criteria (Altman et al., 2001; Moher, Schulz & Altman, 2001). We thus sought to contribute a methodologically sound study of the effects of ICT on early literacy outcomes in Kindergarten reported closely following the CONSORT criteria for RCT trials.

Web-based technology

Most of the programmes described and evaluated by Torgerson and Zhu (2003) are commercially available packages that use CD or videotape as the medium of delivery. In contrast, almost no research exists on *Web-based* technologies. A series of recent studies (Abrami, Savage, Wade, Hipps & Lopez, 2008; Deault, Savage & Abrami, 2008; Savage, Abrami, Hipps & Deault, in press) from the ABRACADABRA literacy research and development programme has sought to fill this important gap in technology research using a free-access Web-based resource. A major advantage of Web-based tools compared with CDs or videotape is that they do not require individual installation and maintenance, often a problem in schools that seldom have comprehensive technical support.

ABRACADABRA is an acronym standing for A Balanced Reading Approach for Canadians Designed to Achieve Best Results for All. This software was developed by the Center for the Study of Learning and Performance, a multi-university research centre, based at Concordia University, Montreal, Canada. The version of ABRACADABRA used in this research can be found at ⟨http://grover.concordia.ca/abra/version1/abracadabra.html⟩.

The version of ABRACADABRA (hereafter, ABRA) used for this study was developed following a year-long pilot study conducted with first graders in 2004–2005.

The activities used in ABRA were modelled after the National Reading Panel's (NRP) analysis of effective reading interventions (NRP, 2000), focusing specifically on letter–sound knowledge, phoneme blending and reading comprehension tasks involving prediction and sequencing.

In the ABRA studies to date, trained research facilitators have worked on basic literacy skills and story reading with small groups of students away from their classroom environment. These students were seen four times a week for between 9 and 12 weeks. As in the current study, previous studies have contrasted differences between a rime-based 'analytic' phonics intervention teaching children words in riming families (e.g. 'cat', 'hat', 'bat') and which encouraged the use of rime-analogy, and a phoneme-based 'synthetic' phonics intervention focusing on the blending and segmenting of phonemes (Johnston & Watson, 2004). Both were compared against regular classroom teaching.

The most recent RCT study by Savage et al. (in press) found significant improvements in letter knowledge in the analytic phonics programme and significant improvements in phonological awareness, listening comprehension and reading comprehension in the synthetic phonics programme at immediate post-test. Significant effects were evident for phonological awareness and reading fluency at a delayed post-test completed 8 months after the intervention, again favouring the synthetic programme. Effect size analyses confirmed that both interventions had a significant impact on literacy at both post-tests (effect sizes across all literacy measures were significantly different from zero at immediate and delayed post-test).

The ABRA studies reported to date have explored the impacts of interventions on typical children in Year 1. The present study thus sought to explore the effects of ABRA on children in Kindergarten classes (broadly equivalent to Reception in the United Kingdom). We also wished to add to a domain of research where there is very little evidence to date: the effectiveness of phonics programmes for children who experience English as an additional language. Studies by Stuart (1999, 2004) explored the impact of synthetic phonics instruction on Sylheti-speaking children of Bangladeshi origin in inner London, UK. Results of the study suggested that these children did indeed benefit from such early interventions. We wish to explore whether effects can be found using ABRA for children from other low-SES communities in a comparable urban social context where many children and their families experience English as an additional language.

Contrasting the effects of synthetic and analytic phonics programmes

There are only four randomised control studies to date contrasting the impacts of analytic and synthetic phonics approaches (three papers identified in a systematic review by Torgerson, Brooks & Hall, 2006 and Savage et al., in press). These studies have found nonsignificant advantages for synthetic over analytic approaches. Torgerson et al. (2006) however also identify the need for more research on this question using RCT studies before drawing any strong conclusions. One purpose of this study was to add a further RCT study to this field.

The qualitatively distinct impacts of synthetic and analytic phonics

All of the studies of analytic and synthetic phonics to date have used standardised measures of reading ability and related skills. Such measures are a good way to identify whether one approach to phonics is quantitatively superior to another. Most literacy researchers have however assumed that the synthetic and analytic phonics teaching leads to *qualitatively* different ways of reading (Duncan, Seymour & Hill, 1997). For example skills in sounding out and blending of grapheme–phoneme correspondences are assumed to result from synthetic phonics interventions and rime-based analogy use is assumed to result from analytic phonics interventions. Thus a major purpose of the present study is to test this widely held assumption of qualitative differences in literacy cognition following synthetic and analytic phonics programmes.

In order to determine whether two different kinds of phonics interventions have qualitatively different effects on the reading strategies used by children, it is necessary to use carefully designed and controlled experimental measures of rime- and phoneme-level phonological awareness and nonword and word decoding measures rather than standardised reading tests. The limited existing research using such tools is thus first described below firstly for phonological awareness and then for reading and decoding.

Synthetic and analytic phonics and phonological awareness

Evidence exists that children benefit from brief rime-based instruction to rapidly improve explicit rime skills (Goswami & East, 2000). Outcomes here were measured in a common unit task where children were required to articulate the shared rimes in words pairs (e.g. 'c*art*' and 'h*eart*') or pairs of words sharing codas (e.g. 'car*t*', 'boa*t*'). In the Goswami and East intervention study, however, only a rime-based programme was implemented, so the *relative* ease and the specificity of the improvements noted in explicit rime awareness versus other subsyllabic units is unknown. Savage, Carless and Stuart (2003) thus evaluated the effects of a rime-based, a phoneme-based and a 'mixed' (rime- and phoneme-based) intervention on phonological onset–rime and phoneme manipulation skills. Results showed greater phonological onset–rime skills in *all* taught intervention groups. It was concluded that there was no simple association between rime- and phoneme-based teaching intervention and the phonological unit used by children following such interventions. However, as this latter study used an onset–rime blending task rather than the common unit task used by Goswami and East, it may be that task differences might explain different results. This issue is therefore revisited in the present study using a common unit task.

Other evidence has added complexity to the idea that children have a particular facility with rimes. Uhry and Ehri (1999) showed that the division of words into onsets and rimes is easier for words with a CVC structure but this does not hold true for shorter words. Uhry and Ehri found that VC words that broke up words within the rime unit (i-ce) were more easily segmented than CV words (s-igh) that respected the onset–rime boundary. Uhry and Ehri (1999) concluded that it is the position of the rime unit in a syllable and not the structure of the rime unit itself that makes it more accessible to children. This finding was replicated later by Guedens and Sandra (2002), who added an important methodological refinement by matching the CV words precisely with the VC words in terms of phonemes used in Dutch (e.g. 'to' as CV stimulus, 'ot' as VC stimulus). The impact of different kinds of intervention on these early phoneme manipulation abilities is unknown, and yet the measure is potentially a subtle index of the relative use of rime- and phoneme-based strategies. This issue is thus explored in the present study.

The specific impacts of synthetic and analytic phonics on reading and decoding

The way that children decode nonsense words with contrasting rime neighbourhood sizes is another way to explore qualitative differences between synthetic and analytic phonics programmes (e.g. Christensen & Bowey, 2005; Deavers, Solity & Kerfoot, 2000; Savage et al., 2003). Savage et al. (2003) therefore developed a nonword reading task in which high-rime neighbourhood (HRn) nonwords (e.g. 'dat' – with many real word rime neighbours) and low-rime neighbourhood (LRn) nonwords (e.g. 'tav' with few real word neighbours) were contrasted. Overall the HRn and LRn stimulus sets contained the same letter–sound correspondences, so any advantage for HRn words could only reflect use of a rime-based decoding strategy. In the Savage et al. (2003) intervention study all intervention groups performed better than the control on this task but there were no other significant effects for either high-rime or low-rime words for either of the analytic or synthetic intervention groups. On the other hand, Deavers et al. (2000) did find that a rime-based strategy influenced nonword reading for a distinct set of stimuli sharing real word analogues (e.g. the nonword 'dalk' pronounced so as to rhyme with 'talk', rather than pronounced via grapheme–phoneme rules). Given the mixed findings to date, this issue is therefore revisited here by measuring HRn and LRn decoding abilities at pre- and post-test.

Finally, it has been argued that a phoneme-based decoding strategy is useful to young readers because many English words cannot be deciphered by analogy to words sharing rimes. For example, 'unique' words such as '*soap*' cannot be read using an orthographic rime analogy strategy because they have no orthographic rime neighbours (Patterson & Morton, 1985). 'Soap' can however be read using grapheme-to-phoneme correspondences. Unique words are thus used as another outcome measure sensitive to different strategic approaches to decoding in the present study.

The purpose of the present study, research question and specific predictions

The primary purpose of the present study was thus to explore the effectiveness of a Web-based literacy programme that delivered two distinct phonics programmes. The research question in this study was: *Does the ABRACADABRA literacy program produce qualitatively different effects for synthetic and analytic phonics interventions on phonological, word and nonword measures?*

Based on the literature it was predicted that the synthetic phonics group would show specific improvements on: (1) blending and segmenting tasks; (2) articulation of shared codas (final consonant); (3) decoding of nonsense words (low-rime neighbours; and (4) reading of unique words compared with the analytic phonics intervention group. The analytic phonics group was predicted to show gains in: (1) blending and segmenting only of CV words (those maintaining the onset–rime units); (2) articulation of shared rime units; and (3) decoding of nonsense words with high-rime neighbours only. No strong predictions were made regarding the overall *quantitative* effect of each intervention on reading abilities (e.g. in the WRAT reading test) given the existing literature to date showing few differences. All children in the interventions should improve equally on letter/sound knowledge tasks as both groups were taught these correspondences.

Method

Design

This study used a pre–post-test experimental intervention design. There was a full random allocation of participants within each classroom to each of two distinct reading interventions, so as to provide control for all extraneous variables such as teaching methods in the classrooms and unmeasured child characteristics. Data were collected over 2 years with two samples in the same school. The research team was given access to one intervention school located in one English language school board in an urban setting. The choice of school was made by senior officials in the participating school board who reported that this school was identified based on the board's perception of highest educational need.

There were five kinds of outcome measures that can distinguish qualitative effects of different phonics interventions: the three phonological awareness tasks (blending, segmenting and common unit tasks) and two decoding tasks (nonword reading and unique word reading tasks), so these are considered primary measures. WRAT reading and letter–sounds knowledge measures were considered secondary measures.

Participants

All children in four Kindergarten classes ($n = 65$) were initially approached to take part in the study. The final sample ($n = 53$) was based solely on whether formal written parental consent to participate was returned. There were no exclusions or other eligibility requirements for sample membership. Response rates for participation were also equal in each of the four classes (16, 16, 16 and 17). No children were excluded or refused to take part at subsequent phases of the study and all pre- and post-tests and both of the interventions themselves all involved the same $n = 53$ children initially identified as eligible to take part in the study. These details of the sample are specified in the CONSORT flow diagram (Figure 1).

A formal power calculation was made using Power and Precision V3 (Borenstein, Rothstein & Cohen, 2001). This showed that with α set at .05, $n = 26$ participants in each cell and one covariate, this design would yield a .80 power to detect a large effect size in analysis of covariance (where a large effect is defined, following Cohen's 1988 effect size metric as $f = .40$). This level of power appeared reasonable given that a previous study using ABRA by Savage et al. (in press) had identified the presence of large effect sizes in standardised measures of phonological awareness.

Parent questionnaires were also distributed to obtain information about language spoken at home, and mother's education level. Of those 53 parents consenting for their child to take part, 47 responses (88.7% of the total) to the additional parent questionnaire were returned. Analysis of these parental questionnaires revealed that 11% spoke only English at home. The majority of children (61%) spoke English plus another language, most often an East Indian language. The parents of 28% of children in this sample spoke no English at home at all.

Data on mother's educational level was coded using a 6-point scale from 0 = *no formal education* through high school completion to 6 = *completed postgraduate degree*. This scale was drawn from the 2006 census categories used by Statistics Canada. Only 45% of the mothers of children in this sample achieved a formal post-high school qualification. Scores for two elementary education categories and two post-secondary education categories were combined as the frequency count for two of these variables was below 5. This yielded four categories. Analysis of these scores against norms for women in Quebec from Statistics Canada (2006 census) showed that the levels of education in this sample were significantly lower than provincial norms, $\chi^2(3) = 8.31$, $p < .05$, confirming the presence of educational disadvantage.

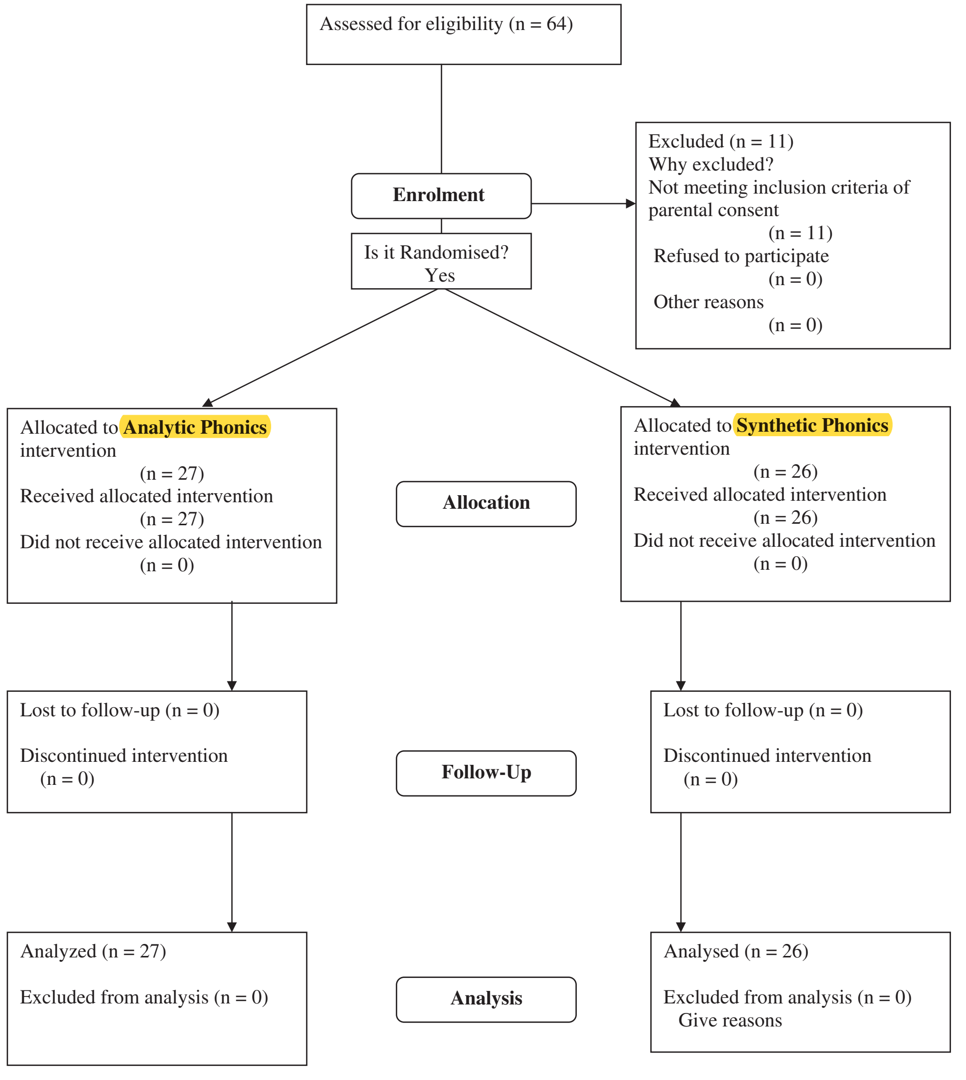

Figure 1. A CONSORT E-flowchart for the ABRACADABRA study.

The children within the two intervention classrooms were randomly assigned to either the synthetic or analytic phonics intervention group using a manual random allocation process (allocation cards pulled blind from a hat) by the first author. There were no restrictions or stratifications in this allocation process. This resulted in a total of $n = 27$ participants in the analytic phonics intervention group and $n = 26$ participants in the synthetic phonics intervention. The first author was also partly responsible for running the interventions as one of the facilitators, and was also responsible for pre- and post-test data collection. As in most educational research where interventions are visible no concealment of treatments was possible. All children in each of the intervention subgroups participated in the study. This sample size allowed the construction of two analytic and two synthetic phonics groups per intervention classroom in each of the 2 years of the study.

Procedure

Pre-testing was conducted at the beginning of the school year and was concluded in 3 weeks. Each child was screened individually by one of two trained facilitators away from their classroom. Following the pre-testing, interventions began immediately in the two assigned classrooms. Interventions were part of the 'centre activities' of the day. This means that the students were engaged in other learning centres and would rotate into the 'ABRA' centre during this designated time period. Small groups of (generally) four children working with ABRA would sit at the designated table in a quiet area in the room with the facilitator. The computer-based program was always the focus for and source of learning activities. Children would engage in ABRA activities around a single computer supported by a facilitator. There was one mouse that rotated from child to child as each took turns.

There were two facilitators, one for each year the study was run. In both cases the facilitator was a trained Master's degree student who was also an experienced early years teacher. The facilitators each received detailed training in the use of ABRA and were highly familiar with and confident in using the system. Training included 2 days of formal training in assessment and intervention, practice time and ongoing weekly follow-up support by the second author and a devoted ABRA project manager. The primary role of the facilitator was to help children to navigate through the appropriate range of activities for each intervention while supporting the socio-academic functions of the group (e.g. maintaining on-task behaviour, observing peers' responses, turn-taking, offering of suggestions, responding appropriately as a group to ABRA prompts such as to blend a word). The facilitator would also judge when the group as a whole had reached mastery level for an activity and should move to the next level of task difficulty.

Every attempt was made to keep the two intervention conditions the same, apart from the different word attack strategies that were taught. Precisely specified lesson plans were developed by the ABRA team and reviewed weekly via reports from facilitators. These features alongside informal observation of interventions ensured high treatment integrity. For this reason, the same facilitator worked with both a synthetic phonics and an analytic phonics group so that teacher characteristics would be consistent. During the session, children not working with ABRA remained in the other centres in the classroom, thereby limiting any cross-contamination of intervention effects. Each small group of four children was generally seen three times a week, depending on school schedule, for 15 minutes. There were sometimes unavoidable delays due to other events taking place in a classroom. Thus it was not always possible to complete three sessions of ABRA every week for all groups. In a minority of cases therefore, the intervention was spread across 16 weeks. In all cases, a total of 40 sessions were completed resulting in 10 hours of instructional time per child.

Interventions for the synthetic and analytic phonics groups followed the same lesson structure, beginning with an Animated Alphabet followed by a 'core activity'. The Animated Alphabet normally took 2–3 minutes allowing 10 minutes to be spent on the core activity. The core activity for each lesson depended on the group. The analytic phonics group's core activities revolved around word families, identifying words that rhymed and manipulating and articulating words at the onset–rime level. The synthetic phonics group's core activities focused on blending and segmenting simple two-phoneme words, identifying words with shared initial and final consonants and forming new words by blending single phonemes. In general, all synthetic phonics groups mastered the CV and VC levels of their core blending activities. All analytic phonics groups showed rime word generation skills. In both interventions, children moved to the next level of difficulty in each activity when they reached 80% mastery.

Following the core activity, each group had a choice activity. The choice activity could be a story option, or a reinforcement game such as Letter Bingo (identifying the correct letter by its name) or letter–sound search (identifying the correct letter by its sound). Groups could also choose to do a different core activity as their choice. When time allowed, the core activity was repeated before ending each session.

General session components

The initial ABRA sessions included general literacy skills activities that were included in both analytic and synthetic phonics interventions with identical frequency and duration. Both groups were read ABRA stories, identified syllables in words and words in sentences and were taught letter–sound correspondences. All children in the classroom used ABRA so students whose parents did not submit consent were allowed to participate in the use of the software, but data were not collected.

Measures

Nine measurements were used in both the pre-screening and post-testing sessions. Most of these were based closely on those used in studies reviewed in the introduction. The PPVT (Peabody Picture Vocabulary Test) was only used at pre-test to assess general receptive verbal ability in the sample.

PPVT vocabulary. The PPVT III (Dunn & Dunn, 1996) is a test of receptive vocabulary. On each item, the examiner reads aloud a target word and children are asked to select the picture that best illustrates the meaning of that target word. The median published split-half reliability of this test is $r = .94$. The mean standard score on the PPVT (Dunn & Dunn, 1996) was 93.11 ($SD = 14.89$), confirming that as a sample the children had slightly below-average vocabulary abilities.

Letter–sound knowledge. In this task the children were shown the 26 letters of the English alphabet in upper-case letters in random order on one page. The children were instructed to say the sound that each letter makes. Words or letter names were not counted as correct answers. The Spearman–Brown internal reliability for this measure was $r = .93$.

Wide Range Achievement Test Word Recognition Subtest. The first 12 words of the Wide Range Achievement Test Word Recognition Subtest (Wilkinson, 1993) were presented and children asked to read them. This approach was used to obtain a brief measure of reading ability suitable for children who were generally non-readers at pre-test. Standard scores could not be collected as norms do not exist for children this young. The Spearman–Brown internal reliability for this task was $r = .84$.

Experimental phonological awareness tasks

Segmenting and blending. This task consisted of a total of 16 monosyllabic words. The same 16 words were used for both the blending portion and the segmenting portion. There were two groups of eight words: consonant–vowel words (e.g. *Sigh*) or vowel–consonant (e.g. *Ice*) words. The words in each group were therefore matched exactly for phonemic similarity.

The tasks were fully counterbalanced during administration. Following a procedure described by Uhry and Ehri (1999), children were shown a prop of a funny character's face, 'Mrs Funnybunny'. Inside the character's mouth were two manipulatives. The participants were introduced to Mrs Funnybunny and then told, 'She likes to play with words and sounds. The facilitator would say, 'If Mrs Funnybunny says /*z*/ /*oo*/ what word could she be saying?' The facilitator would then move the manipulatives with each phoneme (e.g. /*z*/ /*oo*/ for zoo) and then prompt the child to blend those sounds together to figure out what word was trying to be said. When the word was blended the manipulatives would be pushed together to form the whole word. A total of four practice items were included. The Spearman–Brown internal reliability for this task was $r = .86$ for CV words and $r = .72$ for VC words. Using the same materials and procedure, children were asked to break the words down into their phonemic constituents. The Spearman–Brown internal reliability for this task was $r = .84$ for CV words and $r = .84$ for VC words.

Common units tasks. There were two common unit articulation tasks. One required participants to articulate the common 'rime unit' in two words (e.g. '*eet*' in *feet* and *sleet*). The other task required participants to articulate the common coda unit in two words (e.g. '*g*' in *dog* and *bag*). The order of presentation of these tasks was counterbalanced. In each task the children were given four practice items and picture prompts for each of the two words. Practice items could be repeated until the facilitator was sure that the child grasped the nature of the task. The Spearman–Brown internal reliability for the rime articulation was $r = .84$ and for the coda articulation was $r = .82$.

Experimental word and nonword decoding tasks

Nonsense word recognition. This list of 12 words was comprised of two types of stimuli: those words with high-rime neighbours (e.g. *lan*) and those words with low- or no rime neighbours (e.g. *pol*). The words were randomised and presented to the participant on one sheet in lower-case print. The instructions were to 'Name that monster'. The Spearman–Brown internal reliability for this task was $r = .91$.

Unique word recognition. This list consisted of eight words with no rime neighbours with the same spelling (e.g. soap). The children were asked to say them out loud. The Spearman–Brown internal reliability for this task was $r = .86$.

Results

Preliminary data analyses

Preliminary data analysis was first performed to assure that the data set would meet statistical assumptions for multivariate analyses. The pre-test scores revealed three measures that had significant positive skew: WRAT sight vocabulary, nonsense word reading and articulation of rime units. Each of these measures showed patterns consistent with floor effects at pre-test. These variables were thus transformed using an inverse plus constant transformation following the standard procedures described by Tabachnick and Fidell (2001, pp. 82–83). This substantially improved the skew for these four samples.

No other transformations or manipulations of the data were necessary. There were no significant pre-test differences between either of the intervention subgroups in PPVT vocabulary, $F(1, 51) = 3.49$, $p = .07$, *ns*, $\eta_p^2 = .06$.

Main analyses

A general linear model analysis was employed for each measure for each intervention group (synthetic vs analytic) univariate ANOVA. In order to closely control for any pre-test differences between intervention groups, the corresponding pre-test measure was entered as a covariate in each analysis of post-test performance. We sought to test only a limited series of *specific* and theoretically driven predictions that were set out at the end of the Introduction for the phonological awareness, word reading and decoding measures, so conventional significance levels for α ($\alpha = .05$) were used. Bonferonni's corrections were not used as they would tend to be overly conservative in data such as these where measures are highly correlated.

Finally, effect sizes for each analysis of covariance (ANCOVA) run were calculated using partial η^2 (η_p^2) which describes the percentage of variance accounted for by a given variable after controlling for differences at pre-test. Following conventional approaches, a small effect size is $\eta_p^2 = .01$, a medium effect size is $\eta_p^2 > .06$ and a large effect size is $\eta_p^2 > .10$. The obtained β in each ANCOVA for the effect of group is also reported alongside the associated 95% confidence intervals for β. All results are reported in Table 1.

Phonological awareness tasks

Blending and segmenting. The blending CV words task yielded a significant result, $F(1, 50) = 9.05$, $p = .004$, $\eta_p^2 = .15$, with a significant advantage evident in comparisons for the synthetic phonics group over the analytic phonics group. The blending VC words task also yielded a significant result, $F(1, 50) = 4.93$, $p = .04$, $\eta_p^2 = .09$. The advantage here was again for the synthetic phonics intervention. The segmenting task yielded no significant effects either for CV or VC segmentation, $F(1, 50) < .02$, $p = .88$, *ns*, $\eta_p^2 = .000$ for CV words and $F(1, 50) = 3.78$, $p = .06$, *ns*, $\eta_p^2 = .07$ for VC words. This latter effect just escaped significance and showed a medium effect size favouring the synthetic phonics programme. Together then, advantages were evident for phoneme blending, but not segmentation, favouring the group taught synthetic phonics.

Common unit tasks. In the common unit rime articulation task, results showed a significant advantage for the analytic phonics subgroup over the synthetic phonics subgroup, $F(1, 50) = 6.38$, $p = .015$, $\eta_p^2 = .11$. There was also a significant effect for the common unit coda task, $F(1, 50) = 5.64$, $p = .02.$, $\eta_p^2 = .10$, this time favouring the synthetic phonics group. These findings suggest that students who are explicitly taught about rimes are able to identify and articulate that unit. Patterns appear to be specific to rime units in that particular intervention. The same specificity is evident in coda articulation after being taught about individual phonemes.

Word and nonword reading tasks

For the unique word list there was no main effect of intervention group, $F(1, 50) = 0.11$, $p = .74$, *ns*, $\eta_p^2 = .002$. As the HRn and LRn nonwords were explicitly matched on shared grapheme–phoneme correspondences a 2 Intervention Group (synthetic vs analytic vs control) × 2 Nonword Type (HRn vs LRn words) ANCOVA was run with repeated

Table 1. Means, standard deviations, statistical effects and 95% confidence intervals for all outcome measures

Task	Synthetic		Analytic		β of effect and 95% confidence interval for β				P	η_p^2	Effect source
	Pre	Post	Pre	Post	β	Lower bound	Higher bound	F			
Blending CV	1.07 (2.06)	6.27 (2.99)	1.48 (2.64)	3.93 (3.87)	2.61	0.87	4.35	9.05	.004**	.15	S > A
Blending VC	1.96 (2.09)	5.92 (3.02)	2.74 (2.81)	4.59 (3.40)	1.79	0.17	3.41	4.93	.03*	.09	S > A
Segmenting CV	0.88 (2.36)	4.46 (2.82)	1.40 (2.75)	4.63 (3.94)	0.13	– 1.62	1.87	0.02	.88	.00	
Segmenting VC	0.96 (1.97)	5.19 (2.83)	1.89 (3.32)	4.00 (3.65)	1.64	– 0.06	3.32	3.77	.06	.07	
Rime articulation	0.11 (0.59)	0.77 (1.92)	0.67 (2.11)	3.11 (3.72)	– 1.74	– 3.12	– 0.36	6.38	.015**	.11	A > S
Coda articulation	0.96 (2.36)	6.31 (3.52)	2.26 (3.36)	4.70 (3.97)	2.31	0.36	4.25	5.64	.02*	.10	S > A
WRAT word reading	0.62 (1.60)	2.65 (3.06)	0.63 (1.76)	3.22 (4.66)	– 0.54	– 2.07	0.99	0.51	.49	.01	
Unique words	0.56 (1.55)	2.12 (2.42)	0.93 (1.87)	2.15 (2.55)	0.18	– 0.91	1.27	0.11	.74	.00	
High N nonwords	0.04 (0.19)	1.27 (1.66)	0.19 (0.56)	1.22 (1.91)	0.32	– 0.61	1.25	Group × Nonword:		***	
Low N nonwords	0.11 (0.42)	1.12 (1.77)	0.31 (0.71)	1.44 (1.91)	– 0.94	– 1.08	0.89	2.49	.12	.05	
Letter–sounds	7.30 (7.24)	19.61(5.28)	6.22(6.27)	18.81(6.8)	0.26	– 2.59	3.12	0.03	.86	.00	

Notes. $^{*}p < .05$, $^{**}p < 01$. ***As the high-rime neighbourhood and low-rime neighbourhood nonwords are included in an omnibus 2 × 2 analysis, the Group × Nonword effect is the interaction predicted to be significant in research prediction 3, so the interaction effect rather than the main effect of Group is depicted for this effect. S = synthetic phonics intervention; A = analytic phonics intervention.

measures on nonword type and with pre-test nonword type reading score as a covariate. This analysis also revealed no main effect of nonword type, $F(1,49) = 0.16$, $p = .69$, *ns*, $\eta_p^2 < .003$, or group, $F(1,49) = 0.06$, $p = .80$, *ns*, $\eta_p^2 < .001$ and no Intervention Group × Nonword Type interaction effect, $F(1,50) = 2.49$, $p = .12$, *ns*, $\eta_p^2 = .05$. As the HRn and LRn nonwords are included in an omnibus 2 × 2 analysis, the Group × Nonword interaction effect is predicted to be significant in research prediction 3. That is to say, a specific improvement in reading of HRn nonwords and disadvantage in reading LRn nonwords at post-test should have been evident in the analytic phonics condition if children apply rime-based orthographic inferences to reading following rime-based intervention.

Secondary analyses: letter–sound knowledge and WRAT reading

Analysis revealed no significant overall effects for letter–sound knowledge, $F(1,50) = 0.03$, $p = .89$, *ns*, $\eta_p^2 = .001$. There were no differences between each intervention. On the WRAT word reading subtest assessment, no significant intervention group effect was found, $F(1,50) = 0.51$, $p = .49$, *ns*, $\eta_p^2 = .01$.

Discussion

The primary aim of this paper was to answer the question: does the ABRACADABRA literacy programme produce qualitatively different effects for synthetic and analytic phonics interventions on phonological awareness, word and nonword measures? Specific predictions for the synthetic phonics programme were superior blending and segmenting skills for both the VC and the CV words, superior abilities in articulating final consonant codas and superior decoding of all nonsense words irrespective of the number of riming neighbours they have.

The results provided partial support for these predictions as the children in the synthetic phonics subgroup showed specific improvements in both CV and VC phoneme blending tasks compared with children in the analytic phonics group. Children in the synthetic phonics programme also showed a significant and specific advantage for coda articulation in the common unit task and no advantage in the rime common unit tasks.

In contrast, there were no significant advantages for the synthetic phonics group in phoneme segmentation. Why was this? Firstly, there was a discernible trend in post-test improvement in VC segmentation, where a medium-sized effect (η_p^2) of .07 was evident. Such an effect could be practically important but is beyond the statistical power of the present study with $n = 53$ participants to detect as significant. Furthermore phoneme segmentation tasks are among the most difficult phonological tasks (Savage & Carless, 2005; Schatschneider, Fletcher, Francis, Carlson & Foorman, 2004). Finally, children in the synthetic phonics programme here received only around 50 minutes of segmenting training, whereas they received 160 minutes of blending and decoding activity training.

Turning to the analytic phonics programme, this group was predicted to show superior segmentation and blending of CV words where phonemes represent onset and rime and to articulate the shared rime units in words. Similarly, strengths in decoding words with high-rime neighbours were predicted. Results provided only partial support for these predictions. Children showed significant and specific improvements in the common unit rime task. This finding replicates and extends the findings reported by Goswami and East

(2000) to show intervention-specific effects of rime training on rime but not phoneme awareness using comparable tasks to those used by Goswami and East. Against this, there was no advantage in blending or segmenting CV over VC units suggesting that children's rime sensitivity is not reflected in intervention-specific rime blending and segmenting skill, as Savage et al. (2003) reported. This effect might be consistent with the view that the manipulation of CV onset–rime units is different in kind to that of onset–rimes in closed syllables (Guedens & Sandra, 2002; Uhry & Ehri, 1999), possibly reflecting specific perceptual characteristics of CV and VC stimuli.

Turning to the reading and decoding tasks, results here showed that children in the synthetic phonics teaching condition did not read more HRn than LRn nonsense words and read both at comparable levels to the children exposed to the analytic phonics intervention, suggesting instead that they all used grapheme-to-phoneme conversion to decode. This interpretation is also supported by the finding of no qualitative differences in levels of reading for unique words that have no rime neighbours.

These results are thus consistent with some of those reported previously in the literature (e.g. Savage et al., 2003). These results are on first glance not consistent with those reported by Deavers et al. (2000) who reported intervention-specific impacts of analytic phonics programmes. Deavers et al. studied somewhat older children (age 6 years 4 months) who were above-average readers (reading age 6 years 11 months), and who had been exposed to an intensive and daily rime-based phonics programme for nearly 2 academic years. It may be that a tendency to use orthographic rime strategies to pronounce nonsense words only emerges under such circumstances. Why might children in an analytic phonics programme emphasising rimes instead use grapheme-to-phoneme conversion to decode? Reasons for this are unclear, but the rime activities in ABRA may have implicitly directed children to attend to word onsets. It may also be that children preferred to use letter–sound correspondences that they had all been taught in ABRA to aid their early decoding.

There were also no overall advantages for the analytic over the synthetic phonics subgroup in reading sight words from the WRAT. This finding is consistent with and adds to several well-designed recent studies that have contrasted the decoding and reading abilities of young children exposed to synthetic and analytic phonics interventions in older children in Year 1 (Christensen & Bowey, 2005; Savage & Carless, 2005; Torgerson et al., 2006). According to Torgerson et al. (2006) there are only three well-designed RCT studies contrasting synthetic and analytic phonics programmes. They thus describe the evidence base here as 'weak'. This study represents a further contribution on this issue. It reinforces the existing view expressed by Torgerson et al. that there are not marked differences in the impact of synthetic and analytic phonics programmes on reading and decoding levels.

Our study also adds a new element to this literature by indicating that while synthetic and analytic phonics programmes provide quite large and significant early advantage in phoneme and rime phonological awareness, respectively, these qualitative phonological differences do not immediately translate into qualitative differences in the way that children approach word reading or decoding tasks in Kindergarten.

Limitations of the present study

There are several limitations in the present study that need to be considered in evaluating our results. One limitation of the study was that the sample was relatively modest in size.

Nevertheless, the study was sufficiently powerful to detect several large effects for phonological awareness. It is also possible that, as randomisation was not concealed from the researchers who were responsible for both running the study and administering the pre- and post-tests, this may have introduced subtle bias. In practice, it is not possible to disguise the nature of the content of these two interventions. The use of two distinct sub-teams of researchers with one to run the interventions and a second to pre- and post-test the children would however be a methodological improvement in future studies.

The conclusions are also limited to the specific approach used in this study: collaborative interventions in small groups in Kindergarten. This approach may have been effective because children could observe imitate and collaborate with peers. Implementation of ABRA in other ways such as one-to-one tutoring may lead to very different patterns of findings and different patterns may be evident in more experienced or older readers. There were also a limited number of very specific outcome measures in this study, focused as it was on phonics. Furthermore, as the study focused on very young children we were unable to measure reading comprehension growth. Reading comprehension is a key marker of literacy and probably requires children to have both good phonic and verbal comprehension skills.

The research was undertaken in a distinctive and multilingual urban Canadian context. The school was selected by our partner school board as being of particularly high need, so our capacity to generalise to other schools even in this community is unknown. It will therefore be important to replicate the effects in other pedagogical and linguistic contexts.

Another limitation was that the intervention was brief in duration. Our finding of significant effects with a 10-hour intervention is also, in some sense, a strength of the study, and is consistent with findings that brief interventions may be maximally effective (NRP, 2000). Nevertheless the possibility that distinct qualitative patterns might emerge with greater training cannot be ruled out. The current findings that distinct patterns of phonological awareness emerge but are not reflected in qualitative or quantitative differences in decoding are limited to our use of immediate follow-up. Differences between synthetic and analytic programmes might emerge in delayed follow-ups.

Finally, ABRA is also designed to be a tool for teachers. The focus of the present study was on whether technology can be *efficacious*. It is important to add ecological validity to these findings by placing the tool into the hands of well-trained classroom teachers to measure its *effectiveness* in the classroom. Work is currently under way to explore these issues. As ABRA is a free-access Web-based resource we also encourage other researchers and practitioners to take up these challenges with well-executed RCT studies and field trials. The latest version of ABRA consists of 35 graded English instructional activities linked to 17 stories designed to teach emerging readers alphabetics, fluency, comprehension and writing skills and associated subskills. ABRA also contains teacher professional development materials, just in time videos, assessment materials and a communication module (forthcoming) as well as an improved interface (see http://grover.concordia.ca/abra/current/index.php).

Acknowledgements

We would like to thank the staff, parents and children of our partner schools for their willingness to allow us to work with children in their schools for such extended periods of time. We would like to thank Stacy Altman for help in the collection of portions of these data undertaken as partial fulfilment of her MEd thesis. Erin Comaskey collected portions of these data as part of her MA thesis. This study was supported by grants from Fonds québécois de la recherche sur la société et la culture to the second and third author and the Social Sciences and Humanities Research Council of Canada to the second author.

References

Abrami, P.C., Savage, R., Wade, A., Hipps, G. & Lopez, M. (2008). Using technology to assist children learning to read and write. In T. Willoughby & E. Wood (Eds.), *Children's learning in a digital world*. (pp. 129–172). Oxford: Blackwell Publishing.

Altman, D.G., Schulz, K.F., Moher, D., Egger, M., Davidoff, F., Elbourne, D. et al. (2001). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine*, 134, 663–694.

Borenstein, M., Rothstein, H. & Cohen, J. (2001). *Power and precision*. (Vol. 3). Englewood, NJ: BioStat.

Christensen, C.A. & Bowey, J. (2005). The efficacy of orthographic rime, grapheme–phoneme correspondence, and implicit phonics approaches to teaching decoding skills. *Scientific Studies of Reading*, 9, 327–349.

Deault, L., Savage, R.S. & Abrami, P. (2008). *Inattention and response to the ABRACADABRA Web-based literacy intervention*. Manuscript submitted for publication.

Deavers, R., Solity, J. & Kerfoot, J. (2000). The effect of instruction on early nonword reading strategies. *Journal of Research in Reading*, 23, 267–286.

Duncan, L.C., Seymour, P.H.K. & Hill, S. (1997). How important are rhyme and analogy in beginning reading? *Cognition*, 63, 171–208.

Dunn, L.M. & Dunn, L.M. (1996). *Peabody Picture Vocabulary Test*. (3rd edn). Circle Pines, MN: American Guidance Service.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carye, N., Campuzano, L. et al. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Report to Congress, Institute of Educational Sciences National Centre for Educational Evaluation and Regional Assistance.

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

Guedens, A. & Sandra, D. (2002). Beyond implicit phonological knowledge: No support for an onset–rime structure in children's explicit phonological awareness. *Journal of Memory and Language*, 49, 157–182.

Johnston, R.A. & Watson, J.E. (2004). Accelerating the development of reading, spelling and phonemic awareness skills in initial readers. *Reading and Writing: An Interdisciplinary Journal*, 17, 327–357.

Moher, D., Schulz, K.F. & Altman, D.G. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, 357, 1191–1194.

National Reading Panel (NRP) (2000). *Teaching children to read: Reports of the subgroups*. Retrieved 20 November 2008 from http://www.nichd.nih.gov/publications/nrp/report.cfm

Patterson, K.E. & Morton, J. (1985). From orthography to phonology: An attempt at an old interpretation. In K. Patterson, J.C. Marshall & M. Coltheart (Eds.), *Surface dyslexia*. (pp. 335–359). London: Erlbaum.

Savage, R., Abrami, P., Hipps, G. & Deault, L. (in press). A randomised control trial study of the ABRACADABRA reading intervention program in grade 1. *Journal of Educational Psychology*.

Savage, R. & Carless, S. (2005). Learning support assistants can deliver effective reading interventions for 'at-risk' children. *Educational Research*, 47, 45–61.

Savage, R., Carless, S. & Stuart, M. (2003). The effects of rime and phoneme based teaching delivered by learning support assistants. *Journal of Research in Reading*, 26, 211–233.

Schatschneider, C., Fletcher, J.M., Francis, D.J., Carlson, C.D. & Foorman, B. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, 96, 265–282.

Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves reading and spelling in inner-city second language learners. *British Journal of Educational Psychology*, 69, 587–605.

Stuart, M. (2004). Getting ready for reading: A follow-up study of inner city second language learners at the end of Key Stage 1. *British Journal of Educational Psychology*, 74, 15–36.

Tabachnick, B.G. & Fidell, L.S. (2001). *Using multivariate statistics*. (4th edn). New York: Allyn & Bacon.

Torgerson, C., Brooks, G. & Hall, J. (2006). *A systematic review of the research literature on the use of phonics in the teaching of reading and spelling*. Department for Education and Skills Research Rep. No. RR711.

Torgerson, C. & Zhu, D. (2003). A systematic review and meta-analysis of the effectiveness of ICT on literacy learning in English, 5–16. In *Research evidence in education library*. London: EPPI-Centre, Social Sciences

Research Unit Institute of Education. Retrieved 20 November 2008 from http://eppi.ioe.ac.uk/cms/default.aspx?tabid=198

Uhry, J. & Ehri, L. (1999). Ease of segmenting two- and three-phoneme words in Kindergarten: Rime cohesion or vowel salience? *Journal of Educational Psychology*, 91, 594–603.

Wilkinson, G.S. (1993). *The Wide Range Achievement Test*. Wilmington, DE: Wide Range Inc.

Received 25 April 2008; revised version received 19 September 2008.

Address for correspondence: Robert S. Savage, Department of Educational and Counselling Psychology, McGill University, 3700 McTavish Street, Montreal, QC, Canada H3A 1Y2. E-mail: *robert.savage@mcgill.ca*