# Preventative Reading Interventions Teaching Direct Mapping of Graphemes in Texts and Set for Variability aid at–risk learners

**Robert Savage George Georgiou Rauno Parrila & Kristina Maiorino**

## Abstract

We evaluated two experimenter-delivered, small group word reading programs among at-risk poor readers in Grade 1 classes of regular elementary schools using a two-arm dual site matched control trial intervention. At-risk poor word readers ($n$ = 201) were allocated to either a) Direct Mapping and Set-for-Variability (DMSfV); or b) Current or Best-Practices (CBP) small group reading programs, typically for 10-11 hours over 10 weeks. Students were matched on baseline reading and language abilities, parent demographic measures, and on observed regular classroom teaching quality. Results of hierarchical data modeling showed advantages for the DMSfV program ($p$ < .05 for word reading and spelling at post-test and word reading and sentence comprehension at 5-month delayed post-test), with discernible valued added for the DMSfV condition across all follow-up measures. Results support the use of small group preventative literacy intervention models that teach both direct mapping of grapheme-phoneme correspondences in text and set-for-variability.


*Keywords*: reading, intervention, prevention, phonics, dyslexia

# Preventative reading interventions teaching direct mapping of graphemes in texts and set-for-variability aid at-risk learners

The use of evidence-based practices to support at-risk children with low literacy skills is central to their success. While evidence-based practices have given a better start to literacy for many at-risk children, a significant number of children still struggle to read (e.g., Burns et al., 2016; Savage, Carless, & Stuart, 2003; Scammacca et al., 2016). This means, we argue, that researchers should search for optimal novel approaches based on sound theory and careful evidence-based approaches to the evaluation of their impact. We first briefly describe the most effective current practices. We then describe the theory behind our novel intervention before reporting our school-based intervention trial.

## Effective Current Reading Intervention Practices

Meta-analytic reviews of research suggest that well-delivered systematic phonics interventions 'work' to improve literacy (Savage & Cloutier, 2017; Suggate, 2016). However, less is known about what constitutes optimal phonics instruction for poor readers (e.g., McArthur et al., 2012; Torgerson, Brooks, & Hall, 2006). There is evidence that both systematic phonics teaching and shared book reading are important, and that these two practices commonly co-occur in reading intervention studies (Hatcher et al., 1994, 2004, 2006; McArthur et al., 2012; National Reading Panel, 2000). This perspective is also reflected in recent advice to professionals about the importance of 'applying' phonic knowledge (e.g., Common Core Standards, 2015; Education Endowment Fund, 2016). However, the extent to which teaching of grapheme-phoneme correspondences (GPCs) needs to be systematically and explicitly linked to the concurrent reading of authentic texts is unclear from research. Here, we explore the idea that for each phoneme and grapheme–

phoneme correspondence (GPC) taught children should also then read text richly embodying these specific GPC units to consolidate taught skills (e.g., Chen & Savage, 2014). We refer to this practice as 'Direct Mapping' of GPCs. Direct Mapping can be linked to contemporary developmental theorizing wherein increasingly-well specified component GPCs underpin fluent word recognition within a connectionist neural network (Ehri, 2014).

Much research suggests that the combination of phonics and book reading is more effective than phonics delivered in isolation. The 'phonological linkage hypothesis' (Hatcher, Hulme, & Ellis, 1994; Hatcher, Hulme, & Snowling, 2004, Hatcher et al., 2006) expresses the view that children must link graphemes to phonemes and then to text reading to achieve fluency. In a series of intervention studies by Hatcher and colleagues, children were taught GPCs and then read authentic texts in the same intervention sessions. Such approaches proved more effective than teaching phonics in isolation (e.g., Hatcher et al., 1994). While children sometimes wrote the words from texts, most of the time children's attention was not explicitly drawn to the link between specific GPCs taught in any given phonics session and how these units were represented in authentic texts children read. Some suggestive work does exist that has very explicitly linked taught GPCs to text reading. For example, Shapiro and Solity (2008) taught typical Grade 1 children common GPCs and then linked these taught GPCs explicitly to the reading of 'real books' (a range of unselected popular published texts for Grade 1 children selected to contain a high density of taught GPCs). Their results showed this approach improved reading outcomes over regular phonics teaching. However, the experimental and control interventions varied in a number of ways that complicate interpretation (e.g., in use of small groups versus whole

class teaching, massed versus distributed practice, and the teaching of different 'grain sizes' of rimes versus phonemes). This study did not include a delayed post-test to assess retention of learning and possible Hawthorne effects and also explored outcomes in typical readers.

Chen and Savage (2014) ran an RCT intervention with intervention and control conditions and also taught GPCs to Grade 2 children. Children then undertook shared reading of texts with a high density of taught GPCs. This shared book reading always required children to articulate recently taught GPCs – the 'Direct Mapping' of GPCs. Positive effects were again reported for reading outcomes favoring the experimental phonics intervention condition, but this study included no delayed post-test and controls were taught 'word usage,' which was a vocabulary intervention. Thus, we do not yet know that Direct Mapping improves reading over and above the effects of more standard phonics control interventions where GPCs are taught and children then read texts without explicit systematic linkage between the content of the two activities. We thus contrast two interventions: One where the same GPCs taught *on that day* are explicitly and repeatedly reinforced in real texts and a second intervention where phonics teaching and real book reading routinely co-occur but where the linkage between the specific GPCs taught and text reading is not made explicit. Further, we focus specifically on at-risk readers (those scoring below the 30$^{th}$ percentile on a standardized word reading test) and explore effects both at post-test and delayed post-test.

The second issue we address concerns the additional challenges of teaching the complex orthography of English that simultaneously represents both phonological and morphological information (Chomsky, 1970; Venezky, 1999). Little is currently known

about the optimal way to teach the many words that breach the most common phonic rules (though see McArthur et al., 2015, for data on the timing of the teaching of phonic rules and 'exception' words). While the teaching of phonic strategies is extremely important to reading outcomes (Savage, Carless, & Ferraro, 2007), the impact of phonics teaching in English is enhanced by a focus on variable vowel pronunciations likely to be central to development (e.g., Lovett, Lacerenza, Steinbach, & de Palma, 2014; Savage & Stuart, 2001, 2006).

In best-practice synthetic phonics programs children are taught to blend speech sounds (phonemes) associated with letter(s) (graphemes) – for example, 'c'-'a'-'t' to read 'cat' and 'c'-'a'-'tch' to read 'catch'. It is possible to break such decoding skill into component sub-skills (Tunmer & Chapman, 2012). Recent '2-step models of word decoding' explicitly consider a second-step link of word recognition from spelling pronunciation undertaken *after* children have identified and then blended phonemes. This second step requires children having a flexible mental 'set-for-variability' in matching pronunciations derived from GPCs to entries in the mental lexicon that only partly overlap with them. This model acknowledges that going from sounding out of '*w*'..'*aaah*''..'*sssuh*'' to identifying '*was*' is a 'near-magical' feat, which the authors argue is not accounted for at all in current connectionist models of reading (Elbro, de Jong, Houter, & Nielsen, 2012; Tunmer & Chapman, 2012).

These ideas are not wholly new. What Gibson (1965) termed 'set for diversity', later termed 'set for variability' by Venezky (1999), refers to the need for additional processing to navigate spelling-sound inconsistencies. Both Gibson and Venezky explicitly link this strategy to the variation associated with the variable pronunciation of common vowels in

English. This approach has been incorporated into established intervention programs (Foorman, Francis, Davidson, Harm, & Griffin, 2004; Lovett, Lacerenza, & Borden, 2000). Recently, Steacy, Elleman, Lovett, and Compton (2016) isolated the more specific effects of attention to the variable pronunciations of vowels and other taught word reading strategies, showing improved reading of closely related experimental items with variant vowels at post-test compared to a phonics control group that learned about GPCs but did not learn about processing variable vowels. In other works, 'set-for-variability' has been construed as a strategy primarily for correcting the pronunciations of erroneously regularized exception words (e.g., Dyson, Best, Solity, & Hulme, 2017; Zipke, 2016). Dyson et al. conducted a brief experimental intervention with typical readers and found that training in what they term 'mispronunciation correction' with instruction in word meanings facilitated transfer to children's self-correction of untaught experimental exception words. Zipke ran a similar brief intervention with typical readers.

An alternate view of 'set-for-variability' is that it is a process that applies to all words (Elbro & de Jong, 2017; Elbro et al., 2012; Kearns, Rogers, Koriakin, & Al Ghanem, 2016). This broader conception of 'set-for-variability' acknowledges that all pronunciations derived from the application of spelling rules are fundamentally distinct from conventional word pronunciations for any given word. As an example, Kearns et al. note, a /*t*/ phoneme differs in the words *butter*, *cat*, *kitten*, *stand*, *tip*, and *train*. In all cases, the recoded /*t*/ must be resolved with its allophone in pronouncing the correct word. The presence of such phonotactic constraints, as well as schwa-bound phonemes, glottal stop consonants, the co-articulation of phonemes in conventional words, and other issues, led Elbro and de Jong (2017) to develop an explicit model wherein 'set-for-variability' is instrumental in reading

regular and exception words from initial ‘spelling pronunciations’ derived from the application of phonic rules alone. Correlational and longitudinal data also suggest that synthetic phonic blending ability and set-for-variability each explain unique variance in both word and exception word reading (Elbro et al., 2012; Tunmer & Chapman, 2012).

Importantly, none of the existent studies have undertaken a sustained intervention to show that the teaching of a ‘set-for-variability’ strategy with the foundational phonics that are a prerequisite to its use, has a measurable effect on standardized psychometric measures of reading compared to a control intervention that teaches phonics, but does not teach ‘set-for-variability’ to children at-risk of reading difficulties. Given the theorized centrality and generality of ‘set-for-variability’ to word reading above, sustained teaching of this process or strategy should impact psychometric measures of word reading at immediate and delayed post-tests. This distinct approach to teaching and assessing the impact of ‘set-for-variability’ is thus undertaken here.

**The Present Study**

This intervention study examines the impact of distinct and theory-driven reading interventions. The Direct Mapping and Set–for-Variability (DMSfV) intervention incorporates three key features: 1) linking taught grapheme-phoneme correspondences to text containing these specific items (‘Direct Mapping’), 2) an intense focus on teaching alternate vowel digraph pronunciations, and 3) teaching a 2-stage process for reading both regular and exception words (‘Set-for-Variability’). Our experimental hypothesis is that these features improve reading outcomes for at-risk poor readers beyond standard best practices using generic synthetic phonic strategies, rote teaching of high frequency sight words, and generic shared book reading.

# Method

## Design

The study was a dual site cluster matched quasi-experimental efficacy intervention trial that followed a pre-test/post-test/delayed post-test design. We anticipated that as children are often drawn from the same classrooms within schools that there would be dependency in the data. A formal power calculation was thus first undertaken using the Optimal Design software for nested designs (Raudenbush et al., 2011). Analysis of the required number of level-2 (classroom) intervention clusters to achieve conventional power was explored. Modeling revealed that with conventional alpha, and the reasonable assumption that pre-test reading ability as a covariate explains 25% of the variance in post-test reading, $n = 42$ level-2 clusters (classrooms) would be sufficient to detect an effect size of .25 of a standard deviation with conventional power of .80.

Subsequent sampling resulted in the recruitment of 497 children from 42 classrooms nested within 21 schools in 5 school boards. School-level random allocation into one of two intervention arms of the intervention was achieved by first creating a 1, or 2, number code for the two interventions and with the lead researcher then using a true random number generator at www.random.org to allocate schools to conditions. Two additional schools were subsequently added to the sample post-randomization through a process of minimization. This allocation process at the beginning of Grade 1 resulted in 280 students in schools that were allocated to the DMSfV intervention, and 217 allocated to the Current or Best Practices (CBP) intervention.

Interventions were undertaken in the winter semester of Grade 1 following the identification of at-risk participants after a semester of regular instruction in the fall of

Grade 1. This universal screening identified 201 'at-risk' participants (119 in the DMSfV and 82 in the CBP schools). The small group intervention replaced portions of regular classroom language arts teaching about 60-70% of the time. All teachers were observed to use phonic programs as part of their language arts teaching, though varied in the use of these programs. Most often teachers used research-validated programs, most commonly Jolly Phonics ™ (Lloyd & Wernham, 1995), Soundprints ™ (Senecal, 2014), a local analytic word and text-reading program based on research practices, and Success for All ™ (Slavin, 1996).

**Participants**

The 497 participants were drawn from two Canadian provinces (Alberta and Quebec). Following university ethics committee approvals for the study, the students were recruited to the study in two phases in September 2013 and 2014, with the first three authors (i.e., project leaders) overseeing their respective sites in each province. All schools were non-selective non-denominational community elementary schools typical of their urban or suburban contexts, respectively. All students in the kindergarten (Quebec) or Grade 1 (Alberta) classes were initially approached via a parental consent letter. Consent was received when children were in kindergarten or in Grade 1. To achieve consistency and comparability of interventions across schools and conditions, we required that schools allowed us to run small group interventions during regular school time. In one additional site, 9 participants were initially recruited, but it proved impossible to identify agreed upon times for us to undertake small group work with the identified children during regular school time. As a result this site was removed from the study. The final total of 497 students represents the complete sample of children from whom parental informed consents to

participate in the study were received following the guidelines of Canada's tri-council policy on the ethical treatment of human participants.

Girls represented 49.6 % of this sample. The mean age of the sample was 6 years and 4 months (range: five years 6 months to 7 years 4 months). The sample was drawn from urban and suburban schools in each province. Background information collected in a questionnaire revealed that 9.9% of children in the sample were spoken to in a language other than English or French at home. Information was also obtained from this questionnaire on mother's educational levels. Results were compared to Canada's 2011 census information regarding the highest level of education for females aged 25-54. The present sample's maternal education levels were different from those of the national population, $x^2 2(N = 497) = 17.22$, $p < .01$. Our sample consisted of children whose mothers were more likely to hold a university degree and less likely to have a post-school education/diploma or technical training than the national average.

In Quebec, students received classroom instruction with a broadly 50/50 divide between English and French instruction. All of participating families in these schools spoke English at home.

**Materials**

Preliminary analyses showed that up to 60% of the at-risk sample was unable to complete any passage-level comprehension and fluency tests, therefore these measures were not included in the paper.

**Sentence comprehension.** Sentence comprehension was assessed using the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001), a reliable and valid measure (e.g., Fugate, 2003; Waterman, 2003). GRADE was typically administered

to small groups of participants. The sentence comprehension subtest of GRADE requires students to read a sentence and then choose a word that best fits into the context of the sentence from four choices. For example, students were asked to select from among the words ‘jump, ‘put’ ‘guess’, and ‘work’ the word that best fits in the sentence “Can you ____ how many balls are in the box”. A participant’s score was the total number correct. The Spearman-Brown split-half internal reliability in our sample was $r$ = .92.

**Word reading.** Three measures assessed word reading: GRADE (Williams, 2001) Wide Range Achievement Test-4 (WRAT; Wilkinson & Robertson, 2006), and Fry words (Fry, Kress, & Fountoukidis, 2000). In GRADE, students were assessed on word reading and word meaning. These two word-reading sub-tests are combined to produce a ‘Vocabulary’ standard score. In word reading, children identify the word read by the examiner from a choice of four visually and/or phonologically similar words (e.g., the word ‘thank’ from among ‘they’, ‘think’, ‘thank’, ‘take’. Word meaning requires students to read a target word and choose its matching picture from four choices (e.g., to select ‘garbage’ from pictures of a bottle, garbage, flowers, grapes). A participant’s score was the total number correct. The Spearman-Brown split-half internal reliability of the Vocabulary subtest in our sample was $r$ = .83.

In WRAT-4 (Wilkinson & Robertson, 2006, word identification subtest blue form), children were asked to name 15 letters and read from a list of 55 typed lowercase words for a combined word reading maximum score of 70. Discontinuation occurred after 10 consecutive incorrect responses. A participant’s score was the total number correct. The Spearman-Brown split-half internal reliability in our sample was $r$ = .94.

**Fry words.** To further assess the students’ word reading skills, 20 words randomly

selected from the first 200 words in Fry's Instant Word List (Fry, et al., 2000) printed on a sheet of paper were shown to the children one row at a time (5 rows of 4 words each). The students received a point for each word correctly read. The same 20 words were used at pre- and post-test. The Spearman-Brown split-half reliability in our sample was *r* = .86.

**Word Attack.** This task from the Woodcock Johnson III Test of Achievement Form B (Woodcock, et al., 2001) was used to assess students' ability to decode orthographically legal pseudowords of increasing difficulty (e.g., 'hap', 'mel', 'distrum', 'gradly'). The initial items required students to identify the sounds of a few single letters; remaining items required the decoding of increasingly complex letter combinations that follow regular patterns in English orthography, but are pseudowords. Testing was discontinued after six consecutive errors. A participant's score was the total number correct. The Spearman-Brown split-half internal reliability in our sample was *r* = .96.

**Spelling.** In the Woodcock Johnson III Test of Achievement Spelling subtest Form B (Woodcock et al., 2001) children were first asked to write upper or lower case letters (O, T, W, E, R), then to write a lowercase "i", and then asked to write single words to dictation. For each item the word to be written was first said in isolation, and then in a sentence, in which the word was stressed. Finally, the word was said in isolation again. Words were read clearly and repeated when needed, but no other help was provided. Testing was discontinued after six consecutive errors. A participant's score was the total number correct. The Spearman-Brown split half internal reliability in our sample was *r* = .83.

**Phonological skills.** The Phonemic Segmentation Fluency subtest of DIBELS

(Kaminski & Good, 1999) was used to assess students' ability to fluently break three- or four-phoneme words into their individual phonemes in one minute. For example, if the tester read '*ship'*, students had to say /*sh*/ /*i*/ /*p*/. Children were given credit for any correct sounds in the correct order (e.g., responses 'shi'-'p' or 'sh'-'ip' yielded 2 points). The Spearman-Brown split-half internal reliability was $r$ = .96.

**Vocabulary.** Form A of Peabody Picture Vocabulary Test-4 (PPVT; Dunn & Dunn, 2007) was administered to assess vocabulary knowledge. Children pointed to a picture that best corresponded to the word provided by an examiner. A participant's score was the total number correct. The Spearman-Brown split-half internal reliability was $r$ = .94.

**Procedure**

**Test administration.** There were four testing periods: Pre-test (generally September/October of Grade 1), mid-test (January of Grade 1), and post-test (typically May/June of Grade 1) and delayed post-test (generally September/October of Grade 2). The majority of tests were administered in all four phases of testing. The mid-test was a briefer screening tool that served to identify at-risk children. Thus, only the WRAT word reading and Woodcock-Johnson Word Attack measures were administered at mid-test. The DIBELS phoneme segmentation task was administered at the post-test session only. The PPVT was administered in the spring of kindergarten in Quebec and winter Grade 1 in Alberta.

Research Assistants (RAs) conducted all testing. All RAs held or were studying for advanced degrees from education or psychology programs (B.A., M.A. and Ph.D) with some having teaching experience including resource room teaching and additional B.Ed qualifications. The RAs were trained to administer tests by the project coordinator and/or

project leaders. The training sessions lasted approximately 1.5 – 3.5 hours. New RAs accompanied an experienced RA to a school to observe test administration and then were observed administering the tests. Once the assessments were deemed accurately administered by the senior RA, the new RA tested independently.

**Interventions.** Small group interventions were run in the winter semester immediately after the mid-test. All children who performed below the 30$^{th}$ percentile on the WRAT word reading measure at mid-test were judged to be "at risk" and were placed in their corresponding intervention groups. No "at-risk" child was excluded from the intervention. Our sample thus included children with known and/or suspected developmental difficulties, including Cerebral Palsy, Autism, Attention Deficit Hyperactivity Disorder, and Developmental Delay.

Interventions were typically run with groups of 3-4 children outside of the classroom. At-risk children were typically taken out of class for 30 minutes, 3 times a week on days and times agreed upon with the teachers. Group membership only stayed consistent when there was 1 group per class. Otherwise group composition varied depending on class teachers' classroom plans. Occasionally, children from 2 or more classes were assigned to the same group based on availability. Groups typically reflected a range of different reading abilities. If a child missed a lesson, the RA tried to cover missed material with that child during the following lesson. Children received an average of 11-12 hours of small group intervention overall in each of the two conditions over 10 school weeks.

RAs were trained to run the intervention by the researchers and the project manager coordinator or an advanced graduate student who was an experienced resource room

teacher. Students gained skills in assessment and instruction, in the deep structure of English spelling, in the need for multiple strategies to 'decode' this and in practicing supporting phonic and word reading skills. Typically, there was a group meeting of all RAs in each condition lasting approximately two hours. During the workshop the trainer gave an overview of the intervention goals. Then training consisted of reviewing lesson plans, one lesson at a time in detail and acting out certain scenarios that could arise. Typically RAs role-played lessons until all were satisfied they were ready. All RAs could and did contact the project coordinator and/or project leaders at any time with questions, at any point during the intervention. Observers also gave feedback directly following a lesson as part of the treatment integrity process, if required.

***The Direct Mapping and Set-for-Variability Intervention (DMSfV) program*.** The DMSfV program was researcher designed. Further details of the intervention are provided in Appendix 1. Overall the DMSfV approach is characterized by: 1) intense systematic focus especially on variable vowel pronunciations within a synthetic phonics model; 2) very close linkage between grapheme-phonemes taught and shared reading of real books containing a high density of exemplars of that digraph; 3) the teaching of 'set for variability' to map erroneously regularized pronunciations of words to their accurate alternate GPC pronunciation; and 4) significant differentiation of curriculum.

***Common or Best Practices for Word Study (CBP) program*.** The CBP program was also created by this research team. Overall then the CBP approach was characterized by: 1) intense systematic focus especially on blending, and, later, segmenting phonemes within a synthetic phonics model; 2) the daily teaching of common sight words pronunciations and shared book reading, and 3) the *absence* of a) close linkage between

grapheme-phonemes taught and shared reading of real books; b) the *absence* of teaching of 'set for variability' and systematic strategy for variable vowel digraphs. All other non-specific aspects of intervention delivery were comparable to the DMSfV intervention. The CBP Intervention represents quality regular small group literacy instruction and allows an evaluation of whether the *additional* features of DMSFV add value compared to quality instruction. The programs books and their linked GPCS (DMSfV program) and an overview of the content of the two interventions is also provided in Appendix 1.

**Treatment integrity.** In order to assess treatment integrity (TI), both reading interventions were frequently observed. We created a TI rubric that reflected: 1) the specific *Content* of each reading intervention, 2) adherence to the specified *Time Management*, 3) the broader *Teaching Quality*, and 4) broader aspects of the small-group *Learning Environment*. Each component was assessed with a series of 3-5 sub-components on a 3-point scale (0 = 'not done', 1 = 'partly done', 2 = 'fully done'). Observers of TI were all experienced RAs. RAs received training explaining the role and structure of TI in the interventions. Generally, the person doing the TI was also someone who intervened, so they were already trained on the interventions.

Overall, approximately 15% of all of the teaching sessions were observed by RAs. Eight 'live' sessions per intervention were independently observed by two RAs to ensure ongoing inter-rater reliability. Analyses of these scores showed 87.5% agreement in both interventions. Mean scores for each RA were then calculated for all observed sessions separately for each of the four TI components. Mean rankings were uniformly high (ranging from 1.73 to 1.89 on a maximum possible of 2). Mann-Whitney U tests for each TI component by condition (DMSfV vs. CBP) adjusting for multiple contrasts were non-

significant ($p$ >.01 in all cases), confirming that both interventions were equally well implemented.

**Classroom observations.** All regular classes were observed using the Early Literacy and Language Classroom Observation tool (ELLCO K-3; Smith, Brady, & Clark-Chiarelli, 2008) to assess the overall pedagogical quality of the classroom teaching across the two intervention conditions. The ELLCO is a measure of regular early elementary school practices that has been used to assess teaching in other large experimental cluster intervention studies (e.g., Savage et al., 2013). The ELLCO assesses *Classroom Structure and Climate* (CS, classroom climate and management and organization), *Language Environment* (LE, discourse quality and vocabulary learning), *Books and Reading* (BR, resources and phonic, fluency, and comprehension strategy teaching), and *Print and Writing* (PW, writing instruction, environment and products); each component of teaching is assessed on a 5-point scale (1 = Deficient, 2 = Inadequate, 3= Basic, 4 = Strong, and 5 = Exemplary). Training on the ELLCO was given either by a lead researcher or the project coordinator. Training involved careful review of all aspects of the tool, followed by watching a YouTube video of an elementary school English Language Arts (ELA) lesson, which the raters then rated independently. Resulting ELLCO scores were then compared among all raters to ensure at least 80% inter-rater reliability. The process was repeated until the 80% criterion was achieved. Across RAs, the inter-rater agreement prior to observing 'live' sessions was 89.5%. In each classroom, there were always two observers per lesson. Each observer rated the lesson independently and then compared scores and came to an agreed score.

Pairs of RAs observed all English Language Arts lessons twice for approximately 1

hour, or, less occasionally, two 30-minute classes, at times agreed upon with the teacher. Overall, 95% of classes from which at-risk students were drawn were observed. Some teachers did not consent to being observed. In one case the teacher went on leave for the final part of the year. Equal and representative proportions of teachers across the 2 conditions were observed. Analyses of inter-rater reliability of all 'live' class observations showed 93% agreement on CS, 94% agreement on LE, 98% agreement on BR, and 95% agreement on the PW components of ELLCO.

## Results

### Preliminary Data Analyses

RAs scored the tests they administered. A designated RA was responsible for the scoring of all tests, all data entry and calculation of derived scores. All aspects of data scoring was double-checked. Data entry reliability was evaluated on a randomly selected 19.6% of the sample in Quebec and 25 % in Alberta by contrasting the designated RA data sets with that of another trained RA. Data was identical for 99% of items. Negligible differences in randomly distributed errors were not significant ($F < 1$, *n.s.* in all cases).

All data were first screened for the presence of deviations from normality. No significant problems in the data distributions were detected. For the purpose of evaluating the possible impact of outliers, all scores at least two standard deviations above individual variable means were initially considered as potential outliers and removed. Analyses reported below were then conducted for these two databases (with and without candidate outliers) and the results are contrasted. There were no significant differences between results of analyses with and without these candidate outliers, so analyses based on the full sample are presented below. The total missing data across all 3 points of

analysis was 5.02%. The Missing Value Analysis program was run within SPSS, and revealed that missing data were Missing Completely at Random, according to Little's test $x^2$ (31) = 26.611, $p$ = .69, ns. Mortality analyses for the sample missing data at delayed post-test showed no selective loss ($F < 1$, *ns*. in all contrasts). Observation of teachers' practices using the ELLCO observation tool showed that across all intervention conditions, 91% of all observer ratings of teacher practices suggested they were 'Basic', 'Strong' or 'Exemplary'. The latter two categories were evident in nearly 2/3 of all observations.

**Results of the Group Matching Process**

This matched control trial study involved contrasts of two groups of children. Prior to analysis we carefully assessed the quality of match achieved on a comprehensive range of candidate measures including pre-test attainment but also including parent –reported developmental history and the observed quality of their concurrent grade 1 teaching. Results are reported in Table 1. Inspection of these variables show few differences, and only gender reached conventional significance, showing good overall matching.

**Results of the Small Group Intervention: Immediate Post-test.**

This matched control trial used a nested design in which the possible contextual influences on the achievement of the individual participants can be evaluated (e.g., Hox, 2010; Raudenbush & Bryk, 2002). Our data were first analyzed with HLM with classroom as the unit of analysis as our intervention groups within classrooms were not consistent over time. The final HLM models were built in standard 'bottom-up' fashion from preliminary analyses with steps in HLM followed sequentially in order to yield the final models. Model 1 was an Unconditional one-way ANOVA model with random

effects and confirmed that there was classroom level variance at pretest and posttest on attainment measures beyond variance attributable to students (ICC ranged between .09 and .45), and that HLM was appropriate.

An ANCOVA model was appropriate in this design as controls for nested pretest attainment improve the power of such analyses even if the covariate is not statistically significant (e.g., Raudenbush et al., 2011). The final three-level hierarchical model examined whether the significant classroom-level variance on posttest attainment measures (after controlling for School- and Province (provincial site 1 versus site 2)) at level 3, pretest classroom-level attainment variance at level 2, and pre- and posttest student level attainment variance at level 1) was explained by the Condition (DMSfV versus CBP) factor. Equations 1, 2, and 3 describe this final model at the student, classroom, and school levels, or student *i* in classroom *j* in school *k*, respectively.

1) Equation for Student Level 1 Model:

$$Y_{ijk} = \pi_{00k} + \pi_{1jk}\,(\text{PRETEST}) + e_{ijk}$$

2) Equations for Classroom Level 2 Model:

$$\pi_{00k} = \beta_{00k} + \beta_{01k}*(\text{PRETEST ATTAINMENT}) + \beta_{02k}*(\text{INTERVENTION}) + r_{0jk}$$

3) Equations for School and Province Level 3 Model:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(\text{School})_{1k} + \gamma_{002}(\text{Site})_{2k} + \mu_{00k}$$

In these analyses, predictor variables were left uncentered to ease interpretation as ratio-level raw scores have a meaningful zero point value. Identical 3-level models were run for all variables with mid-test WRAT as the covariate.

The means and standard deviations for the two conditions are presented in Table 2. Inspection of this data suggests advantages at post-test for the DMSfV program over CBP at post-tests on WRAT, Woodcock-Johnson, and GRADE reading and spelling measures. Results of these analyses are reported in section A of Table 3 for post-test attainment and in section B of Table 3 for delayed post-test attainment where DMSfV is compared against the zero-ed scores of the CBP condition.

Our main intervention involved phonic work and set-for-variability whereby strategies for lexicalizing phonic skills are taught so children can read words even where they break standard phonic rules. Our primary outcome measures were WRAT and WJII pseudoword reading, reflecting this content and we used unadjusted alpha for these two measures. All other measures were secondary outcomes and alphas adjustments were made for the total number of secondary outcomes ($\alpha = .05/ 5 = .01$ at post-test, $\alpha = .05/ 3 = .017$ at delayed post-test).

The results show that there is a significant effect of condition favoring DMSfV over CBP for WRAT word reading, and Woodcock-Johnson spelling at post-test, and on WRAT word reading and GRADE Sentence comprehension at delayed post-test ($p$ <05). There was also a near-significant effect of condition on Woodcock-Johnson pseudoword reading and GRADE vocabulary at delayed post-test ($p = .056$ in both cases) again favoring DMSfV. Other effects did not reach significance.

**Intervention Effect Size Analyses**

Effect sizes (Hedges' *G*, Hedges, 2007) are also reported in Table 2. Inspection of the effect sizes reveals generally small to medium effects at post and delayed post-tests across all measures generally favoring the DMSfV intervention.

## Discussion

The present study explored the hypothesis that an intervention that included teaching of Direct Mapping of taught GPCs and set-for-variability added measurable value to reading outcomes beyond a current or best practices program that included best practices in the teaching of both phonics and sight word reading. Students were matched on candidate extraneous variables. Preliminary analyses confirmed that there were no effects of age, mother's education, home language, or parent-reported language or literacy difficulties across the 2 intervention groups, nor were there any differences in attainment or in language measures. Direct observation of regular teaching practices suggested students were matched on these as well.

Results from careful contrasts by condition within nested models showed that the DMSfV intervention produced significant advantages in word reading, spelling and sentence comprehension over the CBP program. Analyses of effect sizes showed a similar pattern of advantage for the DMSfV program across all measures that was evident at both post-test and at delayed post-test. To our knowledge, this is the first time that CBP and DMSfV have been contrasted. Previous studies have indicated that components of these programs can be effective. For example, Chen and Savage (2014) incorporated the mapping of taught graphemes when reading texts, and Lovett and colleagues (2014) highlighted variable vowel rules. Here, the aim was to explore the immediate and longer-term impact of this multi-component intervention on standardized literacy outcomes on universally screened at-risk readers, to evaluate the utility of teaching set-for-variability, preceded by instruction in the direct mapping of phonemes. We evaluated whether this approach was superior to current or best practices. Our results supported this view.

How might such results be theorized? It was noted in the introduction that there exist a number of somewhat distinct conceptions of 'set-for-variability'. Both Gibson and Venezky's original work, and some recent intervention research (e.g., Steacy et al., 2016) focuses pedagogical attention on the variable pronunciations of many vowels in English. Others construe 'set-for-variability' as a strategy primarily for correcting the pronunciations of erroneously regularized exception words (e.g., Dyson et al., 2017; Zipke, 2016). Dyson et al. frame their results as consistent with triangle connectionist models of reading (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996), wherein distinct and distributed semantic resources aid orthography-to-phonology processing.

An alternate broader conception of 'set-for-variability' is that it is a process that applies to all words (Elbro & de Jong, 2017; Kearns et al., 2016). From this view, 'set-for-variability' is strongly implicated in the reading of all words because initial 'spelling pronunciations' derived from the application of phonic rules alone are affected by multiple phonotactic constraints and are representations of linear strings of decoded phonemes rather than the commonly co-articulated phonemes in conventional words. Spelling pronunciations are thus seen as distinct from conventional word pronunciations, and set-for-variability is the vehicle for linking the two pronunciations.

Given this centrality of 'set-for-variability' to output representations of all words being learned, we hypothesized that teaching it to young at-risk readers using regular and exception words as content, should have a measurable and sustained impact on standardized psychometric measures of reading. To this end, we used a multicomponent approach. This included oral activities such as 'Simon says touch your…. 'ar'-'m' / 'k'-'n'-'ee'' to teach the component skill of word recognition from given spelling

pronunciations for both regular and exception words, within a clear semantic category (i.e., here, body parts, elsewhere animals, numbers).

As our intervention and outcome measures merged regular and exception words, and we found robust effects at post-test and delayed post-test, our preferred interpretation of these results is that 'set-for-variability' is as a universal feature of the processing of all words in attempts to link spelling pronunciations to conventional pronunciations in learning to read (Elbro & de Jong, 2017). More work is clearly needed to confirm this hypothesis. This broader interpretation is, however, also consistent with other experimental and longitudinal work on 'set-for-variability' wherein effects were evident for regular and exception words (Elbro et al., 2012).

In the broadest theoretical terms, our results suggest that the provision of additional strategies helps children to map graphemes, including complex vowel digraphs, to multiple words in texts within a developmental phase model (Ehri, 2014; Savage & Stuart, 2006). From this perspective, reading development is marked by greater specificity in phonemically underpinned word representations. It is, however, important to note that the DMSfV approach encouraged strategic and meta-cognitive approaches - requiring rule evaluation, the matching of candidate pronunciations to known words in the lexicon, and sometimes the assessment of semantic and syntactic coherence of the products. A comprehensive model of the processes implicated in using 'set-for-variability' to learn to read arguably remains to be elucidated.

**Limitations of the Present Study**

The present study had a number of limitations, most notable amongst these, we lost one measure, the GRADE at delayed post-test and one sub-test at post-test in one site –

so results for this measure, while based on 75% of the at-risk participants, needs to be undertaken with some caution. These losses were, in all cases, dictated by our partner teachers so were unavoidable: They reflect the realities of running large-scale multi-site field–based studies where events occur beyond experimenters' control. Importantly, these limitations did not confound our primary contrasts of DMSfV and CBP on primary outcomes.

It is not possible to say which (if any) specific part of the multi-component DMSfV intervention was more effective, though the finding that effects were strongest for word reading over pseudoword reading may indicate that the set-for-variability aspect of the intervention was important. Future studies might titrate interventions and their impacts, with the understanding that phonemic blending skill is a prerequisite for the use of set-for-variability strategies. In order to further specify the SfV impact, the future studies need to compare SfV and non-SfV conditions with readers who already have the prerequisite phonics skills. Methodologically speaking, teaching multiple pronunciations of graphemes to a control group who are not also taught the SfV strategies might isolate the specific impact of set-for-variability. However, this is problematic with at-risk readers whose phonics skills require attention and who usually require explicit strategy instruction. For this reason, this precise control was not exercised here, where a delayed post-test design also meant a wait list approach to the teaching of set-for-variability for control children was also not possible. Future work using wait list controls might usefully explore this question. In addition, future work would also arguably benefit from explicit contrasts of matched standardized regular and exception word lists as well as the inclusion of experimental measures of transfer from pre- to post-test, and of the extent of

semantic knowledge of words on standardized reading tests used.

We would also argue that ‘set-for-variability’ is by definition a multi-component approach that is predicated upon basic phonemic decoding skills - the other part of our intervention, as well as a host of other abilities such as word knowledge. One reason for taking a multicomponent approach to pedagogy here is that if ‘set-for-variability’ holds potential to be added to best educational practices, and using universal screening approaches now common in schools, we will probably find (as we did here) many children who need support in basic phonics development before they can use set-for-variability. Thus, we have to start where children are developmentally first supporting foundational skills for set-for-variability, as we did here. To this end we taught phonics thoroughly, but we also introduced children to the idea that there is an imperfect match between strings of phonemes and stored words in the lexicon and from there to exemplification of variable vowels and a strategy for using them. As our sample was comprised of children at a very early stage of learning to read, we judged a relatively ‘introductory’ approach to SfV was needed, including only one or two alternate vowel phonemes in our focus on the idea of variability in vowels. The effectiveness of a more extended approach to teaching multiple vowel digraphs in a more comprehensive manner might usefully be the subject of programmatic further study.

More generally, the full sample was of higher maternal education than national averages. Canada is an officially bilingual country, so it is possible that exposure to a second language affected results. However, the amount of teaching time and the ambient levels of additional languages were quite different in the two provincial sites, yet there were no significant effects of ‘site’ on intervention results. Children were selected on the

basis of the WRAT word reading test alone, and had near-average profiles on other literacy measures such as pseudoword decoding. Importantly, however, over half of these children remained relatively weak word readers even on age-appropriate sentence comprehension measures after interventions, necessitating ongoing support. DMSfV, while effective, was no 'inoculation' for them. Overall, there is a need for replication and extension of our findings.

**Conclusions and Implications for Practice and Policy**

Our results suggest that interventions including 'lexicalized' phonic strategies for taught variable vowel rules, and where the taught units are densely represented in texts that are shared that day, are more effective than common best practice interventions containing common research-validated practices (e.g., well-delivered differentiated, preventative, thrice weekly small group synthetic phonics supports, the teaching of sight words and shared text reading). The effects of our DMSfV interventions were still evident 5 months after the intervention finished. Our results are consistent with 'two-process models' of phonics (Elbro et al., 2012; Kearns et al., 2016; Tunmer & Chapman, 2012). The results also suggest that our approach of extended teaching of these two processes may have some ecological validity, with measurable impact on standardized measures on unselected samples of below-average readers in regular classroom settings. DMSfV is perhaps by definition, multi-componential and we cannot isolate specific effects on literacy outcomes. On the other hand, a theory arguing that improvements shown were driven by the DM, not the SfV component of our program, could not explain the non-significant effects found for pseudowords. Hence, we cautiously suggest that SfV, albeit probably in close conjunction with direct mapping of GPCs, improved word

reading.

More generally, our findings have implications for the conceptualization and assessment of the learning barriers that may cause literacy difficulties. The evidence from our interventions here also shows promise in developing and improving intervention practice in schools. Most generally of all, optimal intervention theory, policy, and practice probably hinges on the demonstration of value added of optimal new models over the current or best practices such as we demonstrate here.

## References


Burns, M. K., Maki, K., E., Karich, A.C., Hall, M., McComas, J. J., & Helman, L. (2016). Problem analysis at tier 2: Using data to find the category of the problem. In S. R. Jimerson, M. K. Burns, & A. M. van der Heyden (Eds.), *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd edition), (pp. 293-308). New York: Springer.

Chen, V., & Savage, R. S. (2014). Evidence for a simplicity principle: Teaching common complex grapheme-phonemes improves reading and motivation in at-risk readers. *Journal of Research in Reading, 37,* 196–214. doi:10.1111/1467-9817.12022.

Chomsky, C. (1970). Reading, writing, and phonology. *Harvard Educational Review, 40,* 287–309.

Common Core Standards, (2015). Available: http://www.corestandards.org/ELA-Literacy/RF/1/.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test 4th edition*. Pearson: San Antonio Texas.

Dyson, H., Best, W., Solity, J., & Hulme, C. (2017). Training mispronunciation correction and word meanings improves children's ability to learn to read words. *Scientific Studies of Reading, 21,* 392-407. doi: 10.1080/10888438.2017.1315424

Education Endowment Fund, (2016). *Improving Literacy in Key Stage One*. London: Education Endowment Foundation.

Ehri, L. C. (2014) Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18, 5-21,* doi: 10.1080/10888438.2013.819356

Elbro, C., & de Jong, P. F. (2017). Orthographic learning is verbal learning. The role of spelling pronunciations. In Cain, K., Compton, D., & Parrila, R., (Eds.), *Theories of reading development (in press).* Amsterdam: John Benjamins.

Elbro, C., de Jong, P., Houter, D., & Nielsen, A.-M. (2012). From spelling pronunciation to lexical access: A second step in word decoding. *Scientific Studies of Reading. 16,* 341-359. doi: 10.1080/10888438.2013.8119356

Foorman, B. R., Francis, D., Davidson, K. C., Harm, M. W., & Griffin, J. (2004). Variability in text features in six grade 1 basal reading programs. *Scientific Studies of Reading, 8,* 167-197.

Fry, E. B., Kress, J. E., & Fountoukidis, D. L. (2000). *The reading teacher's book of lists*. Paramus, NY: Prentice Hall.

Fugate, M. H. (2003). Review of the group reading assessment and diagnostic evaluation. In B. S. Plak, J. C. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurements yearbook* (pp. pp. 425–429). Lincoln, Nebraska: The Buros Institute of Mental Measurements.

Gibson, E. J. (1965). Learning to read: Experimental psychologists examine the process by which a fundamental intellectual skill is acquired. *Science, 148*, 1066–1072. doi:10.1126/science.148.3673.1066

Good, R. H., & Kaminski, R. A. (Eds.) (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from: http://dibels.uoregon.edu/

Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. J. (2013). *DIBELS Next*. Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from: https://dibels.org/dibelsnext.

Good, R. H., Kaminski, R.A., Shinn, M., Bratten, J., Shinn, M., Laimon, D., Smith, S., & Flindt, N. (2004). *Technical Adequacy of DIBELS: Results of the Early Childhood Research Institute on measuring growth and development* (Technical Report, No. 7). Eugene, OR: University of Oregon.

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*, 41-57.

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

Hatcher, P. J., Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry, 45*, 338-358.

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research, 32*, 385-410. doi:10.1111/j.1468-2958.2006.00281.x

Hedges, L. V. (2007). Effect sizes in cluster-randomized design. *Journal of Educational and Behavioral Statistics, 32*, 341-370. doi: 10.3102/1076998606298043.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York: Routledge.

Johnston, R. S., & Watson, J. E. (2004). Accelerating the development of reading, spelling and phonemic awareness skills in initial readers. *Reading and Writing*, 17, 327-357.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25*, 215-227.

Kearns, D., Rogers, H. J., Koriakin, T., & Al Ghanem, R. (2016). Semantic and phonological ability to adjust decoding: A unique correlate of word reading skill? *Scientific Studies of Reading, 20*, 455-470. doi:10.1080/10888438.2016.1217865

Lloyd, S. M., & Wernham, S. (1995). *Jolly Phonics Workbook*. London: Jolly Learning.

Lovett, M. W., Lacerenza, L., & Borden, S. L. (2000). Putting struggling readers on the PHAST track: A program to integrate phonological and strategy-based remedial reading instruction and maximise outcomes. *Journal of Learning Disabilities, 33*, 458-467.

Lovett, M. W., Lacerenza, L., Steinbach, K. A., & De Palma, M. (2014). Development and evaluation of a research-based intervention program for children and adolescents with reading risabilities. *Perspectives on Language and Literacy, 40*, 21-29.

McArthur, G., Castles, A., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., & Banales, E. (2012). *Phonics training for English-speaking poor readers*. The Cochrane Library. doi: 10.1002/14651858.CD009115.pub2

McArthur, G., Castles, A., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., & Banales, E. (2015). Sight word and phonics training in children with dyslexia. *Journal of Learning Disabilities, 48*, 391-407. doi: 10.1177/0022219413504996

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

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading; Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115. doi:10.1037/0033-295X.103.1.56

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X. F., Martinez, A., & Bloom, H. (2011). *Optimal design software for multi-level and longitudinal research* (Version 3.01) [Software]. Retrieved from: www. wtgrantfoundation.org.

Savage, R. S., Abrami, P., Piquette-Tomei, N., Wood, E., Sanghera-Sidhu, S., Burgos, G., & Delevaux, G. (2013). A pan-Canadian cluster randomised control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology, 15*, 310-328. doi:10.1037/a0031025.

Savage, R. S., & Carless, S. (2008). The impact of early reading interventions delivered by classroom assistants on attainment at the end of year 2. *British Educational Research Journal, 34*, 363-385. doi: 10.1080/01411920701609315

Savage, R. S., Carless, S., & Ferraro, V. (2007). Predicting curriculum and test performance at age 11 years from pupil background, baseline skills and phonological awareness at age 5 years. *Journal of Child Psychology and Psychiatry, 48*, 732-739. doi: 10.1111/j.1469-7610.2007.01746.x

Savage, R. S., 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.

Savage, R. S., & Cloutier, E. (2017). Early reading interventions: The state of the practice and new directions in building causal theoretical models. In Cain, K., Compton, D., & Parrila, R., (Eds.), *Theories of reading development (pp.* 409-436). Amsterdam: John Benjamins.

Savage, R. S., & Stuart, M. (2001). Orthographic analogies and early reading: Explorations of performance and variation in two transfer tasks. *Reading and Writing, 14*, 571-598. doi: 10.1023/A:1012052631557

Savage, R. S., & Stuart, M. (2006). A developmental model of reading acquisition based upon early scaffolding errors and subsequent vowel inferences. *Educational Psychology, 26*, 33-53. doi: 10.1080/01443410500340983.

Scammacca, N. K., Roberts, G. J., Cho, E., Williams, K. J., Roberts, G., Vaughn, S. R., & Carroll, M. (2016). A century of progress reading interventions for students in Grades 4–12, 1914–2014. *Review of Educational Research, 86*, 756-800. doi:0034654316652942

Senecal, L. (2014). Sound Prints: *Shapes of mind*. Retrieved from: www.shapesofmind.ca

Shapiro, L. R., & Solity, J. (2008). Delivering phonological and phonics training within whole-class teaching. *British Journal of Educational Psychology, 78*, 597-620. doi:1.348/00070998x293850.

Slavin, R. E. (1996). Success for All: A summary of research. *Journal of Education for Students Placed at Risk, 1*, 41-76.

Smith, M. W., Brady, J. P., & Clark-Chiarelli, N. (2008). *Early Language and Literacy Classroom Observation K-3 tool* (Research Edition). Baltimore: Paul H. Brookes.

Steacy, L. M., Elleman, A. M., Lovett, M. W., & Compton, D. L. (2016). Exploring differential effects across two decoding treatments on item-level transfer in children with significant word reading difficulties: A new approach for testing intervention elements. *Scientific Studies of Reading, 20,* 283-295. doi:10.1080/10888438.2016.1178267

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. doi: 10.1348/000709904322848806.

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 9,* 77-96. doi: 10.1177/0022219414528540.

Torgerson, C. J., Brooks, G., & Hall, J. (2006). A systematic review of the research literature on the use of phonics in the teaching of reading and spelling. UK: Department for Education and Skills. Retrieved from: http://www.dfes.gov.uk/research/data/uploadfiles/RR711_.pdf

Tunmer, W. E., & Chapman, J. W. (2012). Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills?. *Scientific Studies of Reading, 16,* 122-140. doi 10.1080/10888438.2010.542527

Venezky, R. L. (1999). *The American way of spelling: The structure and origins of American English orthography*. New York, NY: Guilford Press.

Vousden, J. I., Ellefson, M. R., Solity, J., & Chater, N. (2011). Simplifying reading: Applying the simplicity principle to reading. *Cognitive Science, 35,* 34-78. doi: 10.1111/j.1551-6709.2010.01134.x

Waterman, B. B. (2003). Review of the group reading assessment and diagnostic evaluation. In B. S. Plak, J. C. Impara & R. A. Spies (Eds.), *The fifteenth mental measurements yearbook* (pp. 239–241). Lincoln, Nebraska: The Buros Institute of Mental Measurements.

Wilkinson, G. S., & Robertson G. J. (2006). *Wide Range Achievement Test 4*. PAR: Lutz, Florida.

Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation*: *Teacher's scoring & interpretive manual*. Circle Pines, MN: American Guidance Service.

Woodcock, R. W., McGrew, K.S., & Mather, N. (2001). *Woodcock Johnson III Test of Achievement.* Itasca Illinois: Riverside Publishing.

Zipke, M. (2016). The importance of flexibility of pronunciation in learning to decode: A training study in set for variability. *First Language, 36*, 71-86. doi: 10.11770142723716639495

Table 1: Matching characteristics of the Intervention sample by condition

| Condition | DMSfV | CBP | Significance |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Gender (% female) [b] | 58 | 44 | 3.92 * |
| Chronological Age in months [b] | 76.78 (4.14) | 77.33 (4.47) | -0.92 *n.s.* |
| Parent-reported learning difficulties [c] | 9 | 13 | 0.82 *n.s.* |
| Mother's education [b] | 4.91 (1.37) | 4.69 (1.3) | 0.98 *n.s.* |
| Mother-child language [b] | 1.64 (0.92) | 1.65 (0.97) | -0.09 *n.s.* |
| Father-child language [b] | 1.57 (0.89) | 1.62 (0.88) | -0.17 *n.s.* |
| Peabody Picture Vocabulary Test [a] (kindergarten) | 94.23 (14.41) | 87.87 (16.55) | 3.13 *n.s.* |
| Wide Range Achievement Test reading [a] | 74.96 (10.25) | 74.14 (10.91) | 0.54 *n.s.* |
| GRADE vocabulary composite [a] | 73.58 (12.39) | 71.00 (11.72) | 2.01 *n.s.* |
| Woodcock Johnson III Pseudo-word spelling [a] | 93.88 (15.43) | 94.63 (15.67) | -0.33 *n.s.* |
| Woodcock Johnson III Spelling of sounds [a] | 92.45 (9.54) | 91.69 (12.25) | 0.49 *n.s.* |
| GRADE Listening comprehension [b] | 3.61 (1.74) | 3.51 (1.86) | 0.38 *n.s.* |
| Observer–rated grade 1 teaching [b] | 11.08 (1.47) | 12.53 (3.80) | 2.65 *n.s.* |

*Note*: Values are represented by (a) standard scores, (b) raw scores (c) percentage.

Table 2.

Means and Standard Deviations for the Pre-Test, Post-test and Delayed Post-Test Literacy Measures by Intervention Group

| | DMSfV | | | | CBP | | | | Effect size | Effect size |
|---|---|---|---|---|---|---|---|---|---|---|
| Measure | Pre-test | Mid-test | Post-test | Delay post-test | Pre-test | Mid-test | Post-test | Delay post-test | Post-test | Delayed post-test |
| WRAT [a] | 74.96 (10.25) | 80.55 (9.30) | 97.32 (14.24) | 93.72 (13.47) | 74.14 (10.91) | 79.03 (10.14) | 89.15 (14.44) | 87.93 (13.32) | 0.41 | 0.21 |
| Word attack [a] | 93.88 (15.43) | 100.63 (13.97) | 104.67 (13.82) | 100.83 (12.24) | 94.63 (15.67) | 99.84 (13.13) | 102.23 (13.67) | 96.56 (11.69) | 0.08 | 0.18 |
| Spell [a] | 92.45 (9.54) | | 104.22 (11.73) | 99.21 (11.08) | 91.69 (12.25) | | 99.24 (11.87) | 94.88 (10.56) | 0.27 | 0.21 |
| Vocab [a] | 73.58 (12.39) | | 86.81 (15.72) | 81.95 (16.01) | 71.00 (11.72) | | 81.83 (15.17) | 75.60 (15.86) | 0.28 | 0.30 |
| Sentence comp [b] | | | 6.25 (3.53) | 7.09 (4.03) | | | 6.65 (4.4) | 5.12 (2.41) | -0.01 | 0.30 |
| Fry words [b] | 2.51 (2.65) | | 13.59 (5.49) | | 2.21 (2.45) | | 11.99 (6.2) | | 0.22 | |
| DIBELS PSF [b] | 15.31 (15.51) | | 39.44 (16.77) | | 17.13 (14.81) | | 34.54 (17.55) | | 0.32 | |

*Note*: Values are represented by (a) standard scores, (b) raw scores.
*Key:*

| | |
|---|---|
| WRAT | Wide Range Achievement Test III, Reading sub-test |
| Word attack | Woodcock-Johnson III Test of Achievement, Pseudoword reading sub-test |
| Spelling | Woodcock-Johnson III Test of Achievement, Spelling sub-test |
| Vocab | Group Reading and Diagnostic Evaluation, Vocabulary Composite score |
| Sentence comp | Group Reading and Diagnostic Evaluation, Reading Comprehension Composite score |
| Blending words | Comprehensive, Test of Phonological Processing, Blending Words sub-test |
| Fry words | 20 words from Fry high frequency word list |
| DIBELS PSF | DIBELS Phoneme Segmentation Fluency sub-test |

Table 3
*HLM results for the effect of tier 2 intervention condition on post-test attainment*

| Variable | Classroom-Level Model | | School/Province-Level Model | |
|---|---|---|---|---|
| | Coefficient | *SE* | Coefficient | *SE* |
| A) Post-test | | | | |
| WRAT word reading Post = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.47 | 0.02 * | | |
| Intervention | 2.27 | 0.89 * | | |
| School | | | 0.17 | 0.10 |
| Province | | | 2.64 | 2.23 |
| Woodcock-Johnson Word Attack Post = Dependent Variable | | | | |
| WRAT word reading Mid-test | 0.33 | 0.12 ** | | |
| Intervention | 0.69 | 0.71 | | |
| School | | | 0.96 | 0.08 |
| Province | | | 0.71 | 1.79 |
| Woodcock-Johnson Spelling = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.02 | 0.01 ** | | |
| Intervention | 1.42 | 0.55 * | | |
| School | | | 0.04 | 0.06 |
| Province | | | -0.32 | 1.37 |
| GRADE Vocabulary = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.15 | 0.10 | | |
| Intervention | 3.72 | 1.95 + | | |
| School | | | -0.15 | 0.21 |
| GRADE Sentence comp = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.01 | 0.01 | | |
| Intervention | --0.06 | 0.66 | | |
| School | | | -0.09 | 0.08 |

| | | | | |
|---|---|---|---|---|
| Fry Words = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.04 | 0.02 * | | |
| Intervention | 1.44 | 0.90 | | |
| School | | | 0.05 | 0.09 |
| Province | | | - 2.64 | 2.24 |
| | | | | |
| DIBELS Phonemic Segmentation Fluency = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.21 | 0.14 | | |
| Intervention | 3.05 | 2.67 | | |
| School | | | 0.54 | 0.30 + |
| Province | | | 2.64 | 2.23 |
| | | | | |
| B) Delayed post-test | | | | |
| WRAT word reading = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.04 | 0.02 + | | |
| Intervention | 2.16 | 1.02 * | | |
| School | | | 0.08 | 0.11 |
| Province | | | 1.23 | 2.48 |
| | | | | |
| Woodcock-Johnson Word Attack = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.31 | 0.16 + | | |
| Intervention | 1.66 | 0.86 | | |
| School | | | 0.10 | 0.09 |
| Province | | | 2.97 | 2.13 |
| | | | | |
| Woodcock-Johnson Spelling = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.02 | 0.01 * | | |
| Intervention | 1.23 | 0.59 | | |
| School | | | 0.02 | 0.07 |
| Province | | | -0.93 | 1.44 |
| GRADE Vocabulary = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.25 | 0.13 + | | |
| Intervention | 5.06 | 2.69 | | |

| | | | | |
|---|---|---|---|---|
| School | | | 0.19 | 0.27 |
| | | | | |
| GRADE Sentence comp = Dependent Variable | | | | |
| WRAT word reading mid-test | 0.09 | 0.06 | | |
| Intervention | 1.91 | 0.69 ** | | |
| School | | | 1.00 | 0.07 |

* $p < .05$, ** $p < .01$ + $p < .1$

# Appendix 1

## Intervention Overview

### SUMMARY OF THE TWO INTERVENTIONS

| DMSfV | CBP |
|---|---|
| Aspects that are distinct across interventions | |
| Taught GPCs were 'directly mapped' to texts each day | Taught GPCs were NOT 'directly mapped' to texts each day |
| Taught more than one vowel digraph pronunciation | Taught only one vowel digraph pronunciation |
| Taught SfV as additional to phonic blending and to 'flip' vowels | NOT Taught SfV as additional to phonic blending or to 'flip' vowels |
| NO teaching of sight words | Daily teaching of sight words |
| Aspects that are shared across interventions | |
| **Student characteristics**: initial reading and spelling ability, decoding ability, listening comprehension, vocabulary, demographics (except gender), parent-reported needs | |
| **Methodological match**: sampling of participants, training and support for RAs, duration and frequency of classes and of elements within the intervention (GPCs synthetic phonics, shared book reading) intervention duration, comparability of regular classroom teaching, supplemental intervention, treatment integrity, school-level randomization. | |
| **Pedagogy:** small groups RA-delivery, differentiation by task, progression and monitoring, lively activity–based learning, teaching of GPCs, use of synthetic phonics, shared book reading each session. | |

### DMSfV

All lessons were created to include review (2-5 minutes), teaching (5 minutes), practice activity/game (10 minutes), and shared book reading (10 minutes). Throughout the curriculum, the goals were to teach children letter sounds, common digraphs and what sounds they made, principles of blending grapheme-phonemes into words (first orally

from phonemes, then from graphemes), and common rules for decoding English (i.e., -silent e rule or vowel digraphs rule of the first vowel being a short vowel and the vowel being silent). In the DMSfV program, ‘ch’, ‘sh’, ‘ck’, ‘ee’, ‘ea’, ‘oa’, ‘ou’, ‘oo’, ‘ai’, ‘ay’ as well as the vowel lengthening ‘magic e’ rule were introduced after all singletons were mastered. These graphemes reflected the highest frequency GPCs in children’s books (Chen & Savage, 2014). The intervention thus reflected the practices commonly used in intensive systematic synthetic phonics programs (e.g., Johnston & Watson, 2004; Savage et al., 2013). Each lesson was designed to build on previous knowledge taught, with a constant review/reassessment of the children’s knowledge and ability. If children struggled, more time was devoted to a topic. Lessons were also constructed to allow for heavy differentiation – students who struggled were given easier versions of the same task, while more advanced students were given more complicated work. For example, initially, when we were working with sounds, weaker children identified the first sound, stronger children identified the last sound and the strongest the middle sound. Later, when we worked with blending, very weak children were asked to blend two phoneme words while stronger children blended three to four phoneme words. Weaker children were asked to read easier words, while stronger children were asked to read more challenging words and sentences. Overall we differentiated learning by ordering the skills to-be taught (e.g. singleton then digraph grapheme-phoneme knowledge, blending (initial, final and then medial vowel phonemes, shorter, then longer words), SfV elements) as a hierarchy of difficulty for teaching, and then sought in each session to give tasks where each child at their different current levels might experience 80-90% success. Such visible differentiation also allowed an audit of content in treatment integrity

evaluations. RAs used games to create 'playful' active learning in their lessons. One activity, for example, involved the children running to the letter that made the sound at the beginning/middle/end of the word posted on a wall within the room. When playing games, weaker children were often paired with the RA or a stronger student to give extra supports (e.g., prompting them to sound out words or reminding them of a rule).

In every session, children also applied their knowledge by reading (as needed with RAs' aid) books. Appendix 1 below lists texts used, and their linked graphemes. Texts were scanned and selected where at least one passage existed within the book with at least 1.10 or greater ratio of words with versus words without taught graphemes of the day. As we undertook shared reading and scaffolded the reading of all elements a child could not themselves read, we did not formally 'level' texts, rather we selected books we knew to be popular contemporary children's fiction for grade 1. The shared book reading was based on the principle that whatever unit was taught that day, a children's book that included high density of those specific units was sought out and used (Chen & Savage, 2014). For example, when teaching the digraph "ou" as in 'mouse' we used the popular book - The Gruffalo. The 'ou' digraph appears 6 times in a 52-word passage on the first page of the book, so this paragraph was the initial focus of the shared book reading. This level of density of GPCs in selected paragraphs from texts was typical of the DMSfV intervention. During the shared book reading an RA read, as needed, some of the words with the child but always paused to encourage the child to read the word containing the taught grapheme of the day. The shared reading was also cumulative in that children read words with the taught GPCs from previous days as well as that of the same day, providing ongoing practice of taught GPCs.

Importantly, we first taught the children the more general *concept* that phonic rules do not always yield a clear pronunciation of words and that a second process (a 'strategy') is required to match pronunciation strings against known words. Because the strings are not the same as the word children learned have to fit or match them, consistent with the conceptual model from Elbro et al. (2012). Children thus understood through this discussion and exemplification through modelled and copied phoneme assembly of both regular and exception words, that successful decoding involves two processes: phoneme blending and a subsequent strategy of matching these phoneme strings to find words they know. Another important part of the DMSfV intervention was that children were given additional strategies to decode words that did not follow the most commonly occurring grapheme-phoneme correspondence. A lot of time was directed to teaching two phonemes that went with one digraph (e.g., 'ou' in 'mouth' vs. 'shoulder'). Since we taught both sounds to the children, when they came across a word with that sound and it was read incorrectly, we prompted the children to think of any letters that made more than one sound and to try "flipping" the sound to see if the new word created made more sense. This was explained to children as a staged 5-step plan wherein:

A) Children blend phonemes of a letter string, looking for and applying well-taught phonic rules.
B) Children evaluate whether their first attempt to synthesize a pronunciation yields a word they recognise (evaluated with a self-reflective lexical decision 'Is this a word I know?' initially supported by the RA)
C) If a child's first attempt to synthesize a pronunciation did not yield a word they recognised (evaluated with another self-reflective lexical decision 'Is this

a word I know?'), children then identify and replace the vowel pronunciation with an alternative vowel pronunciation they know

D) Children then synthesise this revised phoneme string

E) Children re-evaluate this blended string using the same reflective lexical decision process as above.

This specific strategy sat within a pedagogical frame wherein having first taught children a rule for vowel digraphs with an *invariant* vowel (e.g. with two vowels side by side, the first one usually makes the long sound and the second one is silent). We then introduced words with vowel digraphs that contradict that rule (e.g. mouse, mouth vs shoulder; tooth versus foot). We routinely worked within semantic categories (body parts, classroom objects, numbers) that we expected would be familiar to the children, where these word classes yielded many regular and exception words (hand, arm, leg, versus knee, eye, stomach), as well as occurring frequently in popular texts we used.

We then used a range of activities where the RA teacher sounded out words that were *both regular and irregular* and children sought to identify the word. Thus, this approach was also supported as a two-process notion (Elbro et al., 2012; Tunmer & Chapman, 2012). For example, in the first instance children were provided with the regularized phonemes of a word (e.g., /*w*/-/æ//*s*//*p*/) and encouraged to think of real words the phoneme sequence could represent. Group-delivered games such as 'Simon says… touch your …(/*n*/*o*/*s*/*e*/, /*ar*/*m*/, /*h*/*a*/*n*/*d*/, /*l*/*e*/*g, sh/ou/l/d/er,* etc), where children were provided with the most common (regular) pronunciation of phonemes in strings were used to animate such learning. Other games used tokens for orally presented letter strings were categorized by children into familiar semantic categories such as body parts,

numbers, furniture etc. Once familiar with the set for variability idea in phonics they both decoded and identified written words. This word-reading task was also closely tied to the reading of real books.

**CBP**

Similar to DMSfV, each lesson was programmatic and sequential, building on previous experience, and structured to contain a review, 10 minutes of phonic work focusing on a number of grapheme to phoneme correspondences each day, and 10 minutes of shared book reading, delivered with the same sense of 'playfulness' as DMSfV. The CBP program focused on blending in the initial sessions. Each phonics session started by practicing the name and sound of common GPCs (usually singletons), followed by an interactive game such as rolling a dice with letters and generating a word using the letters. Children were encouraged to write these words or pseudowords produced. As the intervention progressed more complex words (e.g., CCVC or CVCC structures) were introduced. Segmenting words into phonemes was generally introduced later in the program. As in the DMSfV program, in CBP, 'ch', 'sh', 'ck', 'ee', 'ea', 'oa', 'ou', 'oo', 'ai' 'ay' as well as the vowel lengthening 'magic e' rule were introduced after all singletons were mastered, but there was no systematic teaching of strategies for using variable vowels within words. Where the DMSfV program focused on variant graphemes, the CBP program instead focused more on phoneme blending in consonant clusters such as 'sp-', 'st-'. The CBP program also differed from DMSfV in allocating 7-10 minutes of work to recognizing 'sight words' drawn from the 100 most frequent word list (Vousden, Ellefson, Solity, & Chater, 2011). Sight word reading is a feature in many

descriptions of best practices for intervention (e.g., Wanzek & Kent, 2012). Shared book reading in the CBP program also differed from the DMSfV condition in that books were not chosen to include specific grapheme-to-phoneme units taught on that day. There was heavy differentiation within the CBP program comparable to that in DMSfV.

## Summary of Lessons and Book Characteristics

**DMSfV**

**A. Direct Mapping**

**1 lesson**: Characteristics of books (books have pages, pages have words, words have letters, how to hold a book, authors & illustrators)

**3 lessons**: basic GPCs – matching letters to one sounds. Sample activities would be running to letter flashcards that corresponded to sound said, building words by finding the appropriate letter magnets to match the sounds or backwards hangman where, once given a sentence, children must identify the letters needed to make the words in the sentence.

**4 lessons**: Short vowels. Sample activities included letter dice (2 dice with consonants, 1 with vowels) and having children blend the resulting CVC non word or playing a memory game with CVC words were only the vowel was different (i.e., – cat, cot, cut - pin, pen, pan - pit, pet, pot)

**3 lessons**: Silent e rule. Sample activities would be a memory game or a Go Fish matching a silent e word with its counterpart without the e (i.e., - kit & kite, hat & hate, cub & cube, not & note), blending pseudowords generated by letter dice (where the last dice was either e or blank).
** extra review silent e lessons if needed.

**3 lessons**: Consonant digraphs: sh, ch, ck. Activity was filling in the blanks with the appropriate digraphs or blending words with digraphs

**B. Set-for-Variability**

**5 lessons**: vowel digraphs. Taught the rule that if 2 vowels are side by side the first one is often long while the second is silent. Taught consistent vowel digraphs (e.g., ee, oa). Then taught variations of common vowel digraphs (ea, oo, ou, ai, ay). Gave children strategies to figure out how to pronounce the words. Sample activities were matching words to images they represented, Go Fish with word pairs or sorting words into categories.

**4 lessons**: Consolidation/ Extension / review. Here children were introduced to new words (including prepositions of place & numbers), many of them harder to decode so requiring the use of SfV strategies, and reviewed all previous concepts taught. Sample activities were building sentences with taught words to describe something around the room (i.e.- regularized names of animal pictures hidden in the room & children had to pick out the word cards to describe where the animal was or showing children "monster" faces with irregular amounts of body parts, and children had to select the appropriate word cards to describe the "monster", 'Simon says…' with spelling pronunciations, regularized pronunciations of numbers to be matched against quantities of available objects (e.g. '/t/-/w/-/o/ pencils' etc.

Children spent 10 minutes per session on shared book reading throughout the intervention.

**CBP**

The same structure as above was followed as described in the DM section A above, except that:

1) Taught graphemes were not systematically linked to words in texts during shared book reading.
2) 7-10 minutes per day was devoted to sight word reading of 100 most frequent words from Vousden et al. (2011).

The CBP group did *not* receive section B Set for variability. They learned the most frequent pronunciation of vowels: ee, ea, oo, ou, oa, ai, ay. In the last lessons 4 they were taught to blend the most common CC- structures (sp-, st-, bl-, dr-, br-) and then and –CC items –st, -nd, - ld, -sk, -mp) within CCVC and CVCC structures. Segmenting of these items was also very occasionally undertaken if children could blend all items. Games used to animate tasks were broadly comparable to DMSfV in all cases but never taught direct mapping or set-for-variability.

Children spent 7-10 minutes per session on the Vousden sight words, and 10 minutes on shared book reading throughout the intervention.

**Books for the intervention**:

Duck in a Truck: 288 words
Albourough, J. (1999). *Duck in the Truck*. London, UK: HarperCollins Publishers, Ltd.

- 19 vowels digraphs where the first is long (there were exceptions like "through" & "shout")
  - Feet
  - Croaks x2
  - Sheep x5
  - Jeep
  - Beep
  - Way
  - Sleepy
  - Goat x3
  - Motorboat
  - Rear
  - Boat
  - Wheels
- 34 instances of /ck/
- 39 short /u/
- At least 3 consonant blends/page
  - 73 /th/ in book
  - 34 /ck/ in book
  - 13 /sh/
  - 12 /tr/
  - 12 /st/
  - 6 /wh/
  - 6 /fr/
  - 4 /sl/
  - 4 /ch/

Where's my Mummy?: 408 words
Crimi, C. (2009). *Where's my Mummy?* Somerville, MA: Candlewick Press.

- 5 silent e words in one page of 54 words
- 12 silent e words in the whole book
- 17 /ee/ in whole book
- 8 /ea/, 7 of which make long /e/, 1 as in head
- 23 /oo/, 9 like look, 14 like moon
- 7 /ou/ like cow
- 7 /sh/

From Head to Toe: 208 words
Carle, E. (1997). *From Head to Toe*. New York: Harper Collins.

- 7 instance of double vowels where the first is long
  - shoulders
  - raise
  - monkey
  - seal
  - knees
  - donkey
  - toe

  * exceptions to rule: penguin, head, foot
- 28 short /a/ vowel
- silent e words:
  - wave

* Note, while this book was easy and good initially for short vowels, we used it later because it talked about animal and body parts, which we covered. It is also a good global review and had double vowels that followed the rule, that didn't follow the rule, and a silent e word. It also exposed kids to longer words they had to sound out and blend, like crocodile and buffalo.

Smash, Crash: 400 words
Scieszka, J (2008). *Smash Crash*. New York: Simon and Schuster.

- 33 /sh/ in book (smash, crash, she, shadow, should, smashing, crashing)
- 8 Double vowels were first one is long:
    - Always
    - Road
    - Mean
    - Wheeler
    - Weird
    - Playing
    - Play
    - Speed
    - Cream (x3)

- 9 silent e: made, name, race, time (x2) huge, these (x2), crane

The Gruffalo: 665 words
Donaldson, J. (1999). The Gruffalo. London: Macmillan Children's Books.

- /ou/ = mouse -> 35 times in book & /ou/ = shoulder 5 times in book (plus other words with different sounds like through and poisonous)
- /ee/ = meet -> 19 times in book
- /ea/ = beam -> 13 times in book
- silent e -> 16 times in book
- /oo/ = food -> 9 times in book & /oo/ = good -> 19 times in book

The Gruffalo has several passages that are dense with a specific sound or rule, so even if the density for the whole book is not high, it is for certain passages (i.e., – 9 instances of /ou/ with various pronunciations in a 52 word passage or 6 silent e words in a 58 word passage)

Gruffalo's Child: 607 words
Donaldson, J. (2004). *The Gruffalo's Child*. London: Macmillan Children's Books

- 22 silent e words
- 28 vowel digraphs were the first is long and the second silent, of which
    - /ee/ 7 times
    - /ea/ 12 times

    * /ea/ & /ee/ were taught together as they make the same sound
- /ou/ as in cow 24 times, /ou/ as in shoulder twice and 10 instances of /ou/ as a different sound.

The Gruffalo's Child also several passages that are dense with a specific sound or rule, so even if the density for the whole book is not high, it is for certain passages (i.e., – 6 instances of /ou/ with various pronunciations in a 50 word passage or 3 silent e words in a 34 word passage or 4 instances of /ai/ in a 34 word passage)

We're Going on a Bear Hunt: 410 words
Rosen, M. (1989). *We're Going on a Bear Hunt*. London: Walker Books Ltd.

- 16 /ch/
- 22 /sh/ (of which 10 in a 26 word passage)
- 9 /ck/ (of which 7 in a 52 word passage)
- 14 instances of vowel digraphs where the first is long and the second silent:
    - day x6
    - deep
    - tiptoe x6
    - ears
- oo = moon
    - oozy
    - hoo x 5
    - woo x5
    - gloomy
    - bedroom
- oo = door
    - door x 4

**Additional notes on book usage**

These books were used for more than one lesson and were used to review previous lessons as well as give example of the new GPCs learned. Sometimes books were not, in their entirety, good examples of the sound or rule of that lesson, however they contained at least one passage that contained a higher density of the rule or sound. When this occurred, that passage was focused on, and the rest of the book was read as review.

During shared book reading, words that followed the rules/sounds taught were always emphasized when reading. For variable vowels, in the first lesson we would typically teach a primary pronunciation of /ou/ (e.g., as in house or mouse) and then, when reading the book, the children would be encouraged to read these /ou/ words only. Subsequently, after additional teaching of alternate vowel pronunciations we would get the children to read the words “shoulder” and “boulder” in the Gruffalo’s child, but not other words containing /ou/ that didn’t follow the rule. Then children would be encouraged to read words with both pronunciations (‘shoulder’ & ‘mouse’).