SCIENTIFIC STUDIES OF READING
https://doi.org/10.1080/10888438.2019.1669607

Routledge
Taylor & Francis Group

Check for updates

The Effects of Teaching Complex Grapheme-Phoneme Correspondences: Evidence from a Dual Site Cluster Trial with At-Risk Grade 2 Students

Robert Savage, George Georgiou, Rauno Parrila, Kristina Maiorino, Kristy Dunn, and Giovani Burgos

[a]University College London, London, UK; [b]University of Alberta, Edmonton, Canada; [c]Macquarie University, Sydney, Australia; [d]McGill University, Montreal, Canada; [e]Adelphi University, Garden City, USA

ABSTRACT

We evaluated the impact of teaching complex grapheme–phoneme correspondences (GPC) derived from the Simplicity Principle to at-risk poor readers in Grade 2 classrooms, using a two-arm dual site matched control trial intervention. Poor word readers (n = 149) were allocated to either a) Simplicity GPC (n = 79) or b) Letter-Name Control (n = 70) small group reading programs, and received intervention for 12–15 hours over 12 weeks. Students were matched on baseline reading, language, parent demographics, and observed regular classroom teaching quality. Results of hierarchical data modeling showed advantages for the GPC-group for word reading, pseudoword reading, and sentence comprehension at post-test moderated by pre-test phonological awareness skills. The results provide support for teaching complex GPCs derived from a "Simplicity Principle" as an approach to intervention for word reading, but suggest that children with low PA need additional supports.

The English spelling system is arguably best described as being "quasi-regular" in nature because English is highly inconsistent in the way it maps orthography-phonology relationships (Daniels & Share, 2018; Seidenberg & McClelland, 1989). Many argue that English spelling simultaneously seeks to code both morphemic and phonemic aspects of language (Bowers & Bowers, 2018; Daniels & Share, 2018; Grainger & Ziegler, 2011; Venezky, 1999). Nevertheless, conceptual and computational modeling shows that combinations of 26 letters (graphemes) represent the 44 smallest speech sounds (phonemes) can guide the assembly of word pronunciations from component grapheme-to-phoneme correspondence (GPC) rules (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Gontijo, Gontijo, & Shillcock, 2003), albeit in a complex fashion (Schmalz, Marinus, Coltheart, & Castles, 2015). Gontijo et al. (2003), for example, have argued that approximately 461 GPC rules can be derived from written English. Given these insights, is there value in teaching some of these complex GPCs to struggling readers? We explore this fundamental question here.

Teaching grapheme-phoneme correspondence rules

Phonics instruction involves teaching children to apply GPCs to assemble word pronunciations, often through phoneme blending. Much research suggests that GPC knowledge and phonemic awareness serve as dual foundations of phonics, operating as causal co-requisites to decoding (e.g.,

Q4 **CONTACT** Robert Savage robert.savage@ucl.ac.uk
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hssr.
Supplemental data for this article can be accessed publisher's website.

Byrne, 1999; Hjetland, Brinchmann, Scherer, & Melby-Lervåg, 2017; National Reading Panel, 2000; Schneider, Roth, & Ennemoser, 2000; Stuart & Stainthorp, 2015). Schaaars, Segers, and Verhoeven (2017) followed 73 children at genetic risk of dyslexia and 73 matched controls. In kindergarten the "at risk" children demonstrated weak phonemic awareness. In Grade 1, these same children were less efficient word decoders than controls. Early phonemic awareness and lexical retrieval predicted reading development of both groups of children. Most clearly of all, Hulme, Bowyer-Crane, Carroll, Duff, and Snowling (2012) showed, following an intervention in which teaching of letter-sound knowledge and phonemic awareness improved reading ability, that the improvements in reading were fully mediated by both phonemic awareness and GPC knowledge.

More generally, the evidence for the effectiveness of phonics interventions is also strong. Recently, Savage and Cloutier (2017) provided a tertiary narrative review of nine meta-analytic phonics reviews. All nine reviews showed at least some positive effects for phonics interventions. There is also supportive evidence from large quasi-experimental studies that evaluate national policy shifts toward the use of phonics (Machin & McNally, 2008).

Against this, Suggate (2010) argues that significant effects of phonics interventions are only evident up to Grade 1, with diminishing returns for phonics interventions apparent after Grade 1. There is also a consensus in the field that substantial minorities of "treatment non-responders" continue to struggle to master decoding and word reading even after well-designed and executed phonics interventions (e.g., Jimerson, Burns, & Van Der Heyden, 2016). There is thus much debate currently about the optimal design and duration of sequential phonics programs within popular response-to-intervention (RtI) systems (see Al Otaiba, Kim, Wanzek, Petscher, & Wagner, 2014; Jimerson et al., 2016). Distinct from "phonics" that teaches children how to use GPCs to read and spell words, little attention has been paid to the underlying *content* of phonics programs in terms of the GPCs taught, and more generally still, on principles and theory behind interventions and of "generative" approaches that might potentially lead to more typical development (Compton, Miller, Elleman, & Steacy, 2014; Snowling & Hulme, 2011).

The simplicity principle

Some research has sought to understand the nature and number of GPCs that should be taught to young children (Vousden, Ellefson, Chater, & Solity, 2010; Vousden, Ellefson, Solity, & Chater, 2011). Vousden et al. (2011) generated a database of all words found in 685 popular contemporary children's books read by children aged 5–7 years in the United Kingdom. All GPCs within these words were coded by frequency of occurrence in the children's text. Vousden et al. (2011) used this frequency-coded GPC list, alongside the highest-frequency exception words, to model the optimal explanatory power of GPCs and exception words in predicting the percentage of all words in children's texts rendered readable. Where few GPCs were known, few words could be read. While around 60–70 of the most commonly-occurring GPCs in children's books yielded progressively larger numbers of words that could be read, adding GPCs after this point showed diminishing returns. This optimally efficient type and number of GPC units that lead to greatest generalization in words in children's books is the *Simplicity Principle* for reading.

Vousden et al. (2010), (2011)) undertook statistical modeling of text corpora. Only one published study presents behavioral evidence on the impact of using the Simplicity Principle in children. Chen and Savage (2014) explored the effects of teaching the most frequently-occurring complex GPCs derived from Vousden et al. (2011) on reading of Grade 1 and 2 students from one school using a randomized control trial design. Preliminary analysis showed that the students scored more than two standard deviations below average on a standardized reading test and were collectively an "at-risk" group. In the Simplicity program, children were taught complex GPCs (vowel digraphs and other units such as <a_e>, <pp>, <tch>, <igh>,) ordered by their frequency of occurrence in children's texts. Children then read texts that richly embodied the GPCs taught in that session and spelled target words that included the taught GPCs. Children were not explicitly taught how to blend GPCs. The children in the control condition were exposed to the same content and spelling practice with the exception that they were not exposed to the pronunciation of phonemes associated with graphemes. Thirty-eight students received 30 supplemental small group sessions over nine weeks. Participants in the Simplicity group performed significantly better at post-test than the control group, with generally large effect sizes (Cohen's *d = 1.85, 0.96, & 0.34* for spelling, word recognition, and reading motivation, respectively).

Chen and Savage provide preliminary evidence of the impact of teaching complex GPCs using content derived from the application of the Simplicity Principle compared to an otherwise identical intervention, but where the individual GPCs were not taught. However, the study is limited by the modest and potentially unrepresentative sample. Given strong evidence that decoding is based upon the dual foundation of phonemic awareness and GPC knowledge (Hatcher, Hulme, & Snowling, 2004; Hulme et al., 2012; National Reading Panel, 2000), one could further predict that teaching Simplicity Principle-derived GPCs will be most effective for children with stronger phonological awareness skills. In this sense, individual variation in phonemic awareness may moderate the effects of attempts to create optimal learning environments based on GPC frequency derived from corpus analysis. In Chen and Savage (2004), all students were reasonably capable phoneme blenders and most could blend CVCs; however, blending was not assessed using standardized tasks, limiting comparisons to other studies.

In the current study, we replicated Chen and Savage's study at scale, in a large sample of Grade 2 poor readers drawn from multiple school sites in two distinct locations in Canada. We assessed phonemic awareness with a standardized measure and explored the moderating effects of phonemic awareness on students' response to intervention. Given evidence that successful phonics-based interventions are mediated by the combination of GPC knowledge and phonemic awareness (Hulme et al., 2012), we hypothesized that the improvements in GPC teaching alone would be strongest among children with stronger pre-test phonemic awareness. Hence, we predicted an interaction effect: Pre-test phonemic awareness will moderate literacy outcomes for the Simplicity-based GPC intervention condition. We also reasoned that children with stronger GPC knowledge and phonemic awareness will have a non-linear advantage over peers with weaker skills in these domains because this dual foundation underpins decoding. Decoding is generative in alphabetic spelling systems through the multiple self-teaching opportunities it engenders (Compton et al., 2011). We thus modeled an intervention x phonemic awareness interaction effect. Finally, given that existing theorizing suggests phonemic awareness and GPCs *cause* improvement in reading interventions (Hulme et al., 2012), these effects should still hold when we keep pre-test reading ability constant. We thus modeled that specific prediction here.

Method

Design

The study was a dual site cluster matched quasi-experimental pre-test/post-test efficacy intervention trial.

Participants

The participants of this study were 510 children (255 boys, 255 girls) attending Grade 2 in two Canadian provinces (Alberta and Quebec) recruited following university ethics committee approvals for the study and formal parental consent. The children were initially recruited for a larger study on Response to Intervention covering Kindergarten to Grade 3 (Authors, 2018). Children attended 44 classrooms within 21 schools in five school boards. Detailed information on how the initial sample was recruited can be found in Authors (2018).

We first matched schools for location (urban versus suburban schools) and then allocated 19 schools randomly to the two intervention conditions with 9 and 10 schools in each condition. Two schools consented to take part after randomization, so were subsequently added to the sample through a process of minimization, one being added to each condition. Once the schools were allocated to the two intervention conditions, we tested all 510 children on the Wide Range Achievement Test Word Reading Task (Wilkinson & Robertson, 2006) to identify those at risk for reading difficulties. One hundred forty-nine children (78 boys, 71 girls) with a percentile score equal to or lower than 40 on word reading were deemed “at risk”. No “at-risk” children were excluded from the intervention. The 40th percentile has been widely used as a threshold for reading “at grade level” statewide in the Unites States, including RtI initiatives and in tools such as DIBELS (Jimerson et al., 2016; Mellard & Johnson, 2008), and worldwide (e.g., in Australia; Reynolds, Wheldall, & Madelaine, 2011). Children attending schools in the Simplicity GPC condition received the Simplicity GPC intervention and children attending schools in the Control condition received the Letter Name intervention (see below). Interventions took place in the winter semester of Grade 2. Around 60 to 70% of the time when children participated in the intervention, it was during Language Arts.

A background information questionnaire revealed that 17.4% of mothers and 19.5% of fathers spoke to their child in French and 5.0% of the children in this sample were spoken to in a language other than English or French at home. Information on mother’s educational level from this questionnaire was compared to 2016 Canadian census data on education for females aged 25–54. The screening sample’s maternal University education level was not significantly different from that of the national population, χ^2 $1(N = 510) = 0.19$, $p = .81$, *ns*. Parent-reported data on children’s learning difficulties was also collected and is reported with other matching data in Table 1.

Materials

Preliminary analyses in one site showed that up to 50% of the at-risk sample was unable to complete passage-level comprehension and fluency tests. Unfortunately, in the other site, teachers requested that the passage comprehension tests not be administered. Therefore, these measures were not included in the paper.

Listening comprehension

Listening comprehension was assessed using the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001). This subtest required students to listen while a research

Table 1. Matching characteristics of the intervention sample by condition.

Variable	Simplicity	Control	Significance F (1, 148)
Gender (% female)[c]	53	41	0.15
Parent-reported learning difficulties[c]	7	11	0.24
Chronological Age in months[b]	89.60 (4.78)	89.20 (4.31)	−0.54
Mother’s education[b]	4.78 (1.42)	4.52 (1.22)	0.27
Mother-child language[b]	1.65 (0.91)	1.54 (0.89)	0.11
Father-child language[b]	1.60 (0.89)	1.50 (0.83)	0.60
Peabody Picture Vocabulary Test[a]	91.85 (18.58)	91.07 (16.11	0.05
Wide Range Achievement Test reading[a]	85.04 (11.24)	86.88 (13.17)	−0.91
GRADE vocabulary composite[a]	73.62 (13.09)	73.43 (13.34)	0.94
Woodcock Johnson III Pseudo-word spelling [a]	94.45 (15.19)	96.15 (11.31)	−0.76.
Woodcock Johnson III Spelling of sounds [a]	92.90 (14.70)	93.75 (10.72)	−0.40
GRADE Listening comprehension [b]	3.44 (1.63)	3.07 (1.49)	1.23
Observer–rated grade 2 teaching [b]	10.07 (1.82)	11.20 (1.91)	−1.46

Note. Values are represented by (a) standard scores (b) observed scores (c) percentage. The Province PPVT data were collected at different times (K in Province 1 and Grade 2 in Province 2). We ran an ANOVA by Province and Condition with no effects for “Province” or Province x Condition (all $Fs < 1$).

assistant read aloud a sentence or pair of sentences such as "Neil will cross the street when the signal light changes". The students then answered by marking an appropriate item among four line drawing pictures in a student booklet. Of the test questions, 12 of 17 were single-sentence stimuli and the rest contained two sentences. A participant's score was the total number correct. The Spearman-Brown split-half internal reliability in our sample was .80.

Sentence comprehension

Sentence comprehension was assessed using the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001). This subtest required 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 "wish", "show" "give", and "thank" the word that best fits in the sentence "Remember to ____ your friend for the surprise". A participant's score was the total number correct. The Spearman-Brown split-half internal reliability in our sample was .90.

Word reading

Two measures assessed word reading: GRADE (Williams, 2001) and Wide Range Achievement Test-4 (WRAT; Wilkinson & Robertson, 2006). In GRADE, students were assessed on word reading and word meaning. These two word-reading sub-tests were combined to produce a composite that the GRADE manual labels a "Vocabulary" standard score. In word reading, children identified 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 required 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). The Spearman-Brown split-half internal reliability of the Vocabulary score in our sample was .94.

In WRAT-4 (Wilkinson & Robertson, 2006, word identification subtest blue form), children were first asked to name 15 letters and then read from a list of 55 lowercase words arranged in increasing difficulty. The test was discontinued after 10 consecutive errors and a participant's score was the total number correct (max = 70). The Spearman-Brown split-half internal reliability in our sample was .93.

Word attack

This task from the Woodcock Johnson III Test of Achievement Form B (Woodcock, McGrew, & Mather, 2001) was used to assess students' ability to decode orthographically legal pseudowords of increasing difficulty (e.g., "hap", "mel", "distrum", "gradly"). 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 pseudo-words. A child's score was the total number of correctly read pseudowords (max = 32). The test was discontinued after six consecutive errors. The Spearman-Brown split-half internal reliability in our sample was .97.

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 and then asked to write single words to dictation. The test was discontinued after six consecutive errors and a participant's score was the total number of correctly written letters and words. The Spearman-Brown split half internal reliability in our sample was .92.

Phonemic awareness

The Blending Words subtest of the Comprehensive test of Phonological Processing 2 (CTOPP-2; Wagner, Torgesen, Rashotte, & Pearson, 2013) was used to assess students' ability to blend sounds to words. For example, the tester said 'what word do these sounds make?*/m//ōō//n/*, students had to say "moon". The test was discontinued after three consecutive errors. A child's score was the total number correct (max = 21). The Spearman-Brown split-half internal reliability in our sample was .83.

Receptive 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 the examiner. A participant's score was the total number correct. The test was discontinued after 8 errors in a block of 12 items. The Spearman-Brown split-half internal reliability in our sample was .93.

Procedure

Test administration

There were two testing periods: Pre-test (December/January of Grade 2) and post-test (May/June of Grade 2). At pre-test, WRAT Word Reading and Woodcock-Johnson Word Attack served to identify the at-risk children. Blending Words was also administered at the pre-test session. PPVT was administered in the spring of kindergarten in Quebec and winter of Grade 1 in Alberta and preliminary GRADE reading measures used to assess the quality of the match of the treatment and control groups were collected for all children initially sampled at the start of Grade 1, as part of our larger project (see Authors, 2018).

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. RAs were trained to administer the tests by the project coordinator and/or project leaders. The training sessions lasted approximately 1.5– 3.5 hours. New RAs tested children independently only after being observed by the senior RAs.

The RAs were responsible for the scoring of all tests they administered. A senior RA was responsible for all data entry, and calculation of derived scores. All data scoring was double-checked. An earlier data entry reliability analysis reported in a previous comparable study (Authors, 2018) established that the same RAs achieved 99% inter-rater reliability, with negligible differences in randomly distributed errors being not significant ($F < 1$, *ns*, in all cases).

Interventions

Small group interventions were run in the winter semester immediately after the pre-test. Interventions were typically run with groups of 3–4 children outside of the classroom 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 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. Children received an average of 12–15 hours of small group intervention overall in each of the two conditions over 12 school weeks.

RAs were trained to run the intervention by the researchers and the project coordinator or an advanced graduate student who was an experienced resource room teacher. Students gained skills in assessment and instruction in the details of Simplicity and Control interventions, respectively. Typically, there was a single group meeting of all RAs in each condition lasting approximately two hours. In that meeting, the trainer gave an overview of the intervention goals, reviewed lesson plans one lesson at a time in detail and acted out certain scenarios that could arise. Typically, RAs role-played lessons until they all felt ready. All RAs could and did contact the project coordinator and/or project leaders with questions at any point during the intervention. Observers also gave feedback directly following a lesson as part of the treatment integrity process (see below), if required.

The simplicity GPC intervention program. The Simplicity program was researcher-designed. Throughout the curriculum, the goals were: a) to teach children the most frequently occurring complex GPCs in the order of their frequency in books, b) have children recognize those GPCs in text, and c) give children experience writing those GPCs embedded in words. The Simplicity approach is characterized by a systematic approach to the delivery of teaching of explicitly articulated complex grapheme-phoneme correspondences wherein the order of delivery was based on the identified GPC frequency in authentic children texts. Unlike in Chen and Savage (2014), in the present study, Canadian data were used to construct a database of complex GPCs. Electronic book records of the most frequently issued children's books (juvenile fiction, nonfiction and "picture books") were obtained from the Toronto District Metropolitan Library system between January 1, 2013 and March 30, 2014. Data from issuing patterns of some 100 libraries in the Toronto area were recorded in this database. Using these data, 363 of 500 books were obtained and all main text therein was entered into a word database containing 8636 word types and 179,678 word tokens. This database was then analyzed by Dr. Jonathon Solity's UK team using the procedure described in Vousden et al. (2010), 2011) to create a Canadian-text based Simplicity Principle-derived GPC list. Isolated GPCs were used with the exception of <qu>, which was kept on frequency of co-occurrence and <-ing> because this unit also appears frequently as the gerund form.

Teaching typically involved four steps: 1) Introduction, definition, and spelling of a new "word of the day"; 2) search for that word in authentic children's books selected to represent these words; 3) shared reading of researcher-written texts that repeated the word of the day wherein children and RA co-read the text with children reading the word of the day; and 4) introduction of the "sound of the day" – a grapheme-phoneme correspondence within the word of the day explicitly articulated by the RA. Children said and wrote this grapheme and then identified this grapheme in texts that were often researcher-written and designed to incorporate a high density of taught GPCs, and also wrote a sentence using this grapheme pattern in their notebook, with emphasis on the sound of this GPC. Researcher-written texts sometimes included words containing variants to pronunciations (e.g., unvoiced <s> in texts where the taught GPC was the voiced <s> rule). Such "incidental" variants were not explicitly taught or drawn attention to. The intervention was delivered with significant differentiation of the curriculum. For example, if the GPC of the day was/ʃ/associated with, where an RA judged that child was a weaker reader they would be asked to read the word "she" while a student the RA judged to be a stronger reader might be asked to read "seashell". Stronger students were also encouraged to read whole sentences/pages of the book/short story while weaker students only were asked to read the word of the day or words with the GPC. Stronger students wrote more complex sentences (sometimes with two words containing the GPC) while weaker students only had to write a simple word. During the writing of the sentence the RAs called attention to aspects of writing, such as punctuation and capitalization.

The control intervention program. The Letter-Name control program was also created by this research team. It mirrored the Simplicity GPC intervention in all respects except that children were introduced to the "special spelling" of the day. Here, the RAs did not call specific attention to the phoneme associated with a grapheme; instead, children were taught to identify a grapheme without reference to its corresponding phoneme. RAs referred only to the names of letters constituting complex GPCs to draw children's attention to the relevant units. In all other respects the taught control condition was identical to the GPC intervention condition. Thus, the Control intervention acted as control for all aspects of the Simplicity GPC intervention beyond the instruction in grapheme-phoneme correspondences. An overview of all of the content of the two interventions is also provided in the Appendix.

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 between 3 and 11 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 20% of all of the teaching sessions were observed by RAs. Of these sessions, 10% of the intervention sessions were independently observed by two RAs to ensure ongoing inter-rater reliability. Observations in one provincial site were audio recorded and independently assessed by the first and fourth author. Analyses of all these scores showed 97% agreement in Simplicity and 98% in Control interventions. Mean scores for each RA were then calculated for all observed sessions separately for each of the four TI components. Mean rankings were high (Mean = 1.81, *SD* = 0.21, for Simplicity and 1.67, *SD* = 0.23, for Control on a maximum possible of 2). Mann-Whitney *U* tests for each TI component by condition (Simplicity versus Control), adjusting for multiple contrasts were non-significant ($p > .10$ in all cases), confirming that both interventions were equally well implemented.

Classroom observations

The overall pedagogical quality of all regular classes was observed using the Early Literacy and Language Classroom Observation tool (ELLCO K-3; Smith, Brady, & Clark-Chiarelli, 2008). The ELLCO has been used in other cluster intervention studies to assess teaching quality (e.g., Savage et al., 2013; Savage, Georgiou, Parrila, & Maiorino, 2018). The ELLCO assesses *Classroom Structure and Climate* (CS, classroom climate and management and organization), *Curriculum* (CR, integration of language and literacy, independent learning, diversity), *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 sub-component of teaching is assessed on a 5-point scale (1 = Deficient, 2 = Inadequate, 3 = Basic, 4 = Strong, and 5 = Exemplary). ELLCO training was given either by a lead researcher or the project coordinator. Training involved systematic review of the tool followed by independent rating of a YouTube video of an elementary school English Language Arts (ELA) lesson. Independent ELLCO ratings were then compared and the process repeated where necessary until 80% inter-rater reliability was achieved. Across pairs of RAs, the inter-rater agreement prior to observing "live" sessions was in excess of 90%. In "live" classroom sessions a pair of observers 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 in total at times agreed upon with the teacher. Overall, 70% of classes from which at-risk students were drawn were observed. Equal and representative proportions of teachers across the 2 conditions were observed. Analyses of all class observations showed 99.5% inter-rater agreement on CS, 99% agreement on CR, 98% agreement on LE, 94% agreement on BR, and 93% agreement on the PW components of ELLCO. These observations also suggested that it was very rare for teachers to teach any GPCs in Grade 2.

Results

Preliminary data analyses

All data were first screened for the presence of deviations from normality. No significant problems in the distributions of the measures were detected. For the purpose of evaluating the possible impact of outliers, all scores at least 1.5 and 2 standard deviations above or below a given variable's mean were initially considered as potential outliers. Analyses reported below were then conducted both with and without candidate outliers and the results were contrasted. There were no significant differences between results of analyses with and without the outliers, so analyses based on the full sample are presented below. The total missing data was 1.83%. A Missing Value Analysis revealed that missing data were Missing Completely at Random, according to Little's test x^2 (8) = 4.09, p = .85, *ns*. Observation of teachers' practices using the ELLCO observation tool showed that across all intervention conditions, 99% of all-agreed observer ratings of teacher practices suggested they were "Basic", "Strong" or "Exemplary".

Results of the group matching process

Prior to conducting any analyses, we assessed the quality of match achieved across the two intervention conditions on a comprehensive range of candidate measures including pre-test attainment, parent–reported developmental history, listening comprehension and the observed quality of their concurrent Grade 2 teaching. Results are reported in Table 1. None of the group comparisons reached significance, showing good overall matching by condition.

Intervention results

Raw score data were analyzed with HLM to account for the nested nature of our data (e.g., Hox, 2010; Raudenbush & Bryk, 2002). In standard "bottom-up" fashion, the final HLM models were built sequentially from preliminary analyses. Model 1, an unconditional one-way ANOVA model with random effects, confirmed that there was significant classroom and school-level variance at pre-test and post-test on primary achievement measures and spelling beyond variance attributable to students (intra-class correlations ranged between .05 and .07), and thus that HLM was appropriate. For the GRADE Sentence Comprehension and Vocabulary sub-tests, classroom- and school-level ICC did not exceed .02 at either level.

With the intervention randomized at the school level, the final three-level hierarchical model examined whether variance on post-test achievement outcome measures at school level (level 3) was explained by an interaction between Condition (Simplicity versus Control, at level 3) and children's phonological awareness skill (leve1 1), after controlling for Province (level 3) and pre-test classroom-level achievement (level 2).

An ANCOVA model was appropriate because controls for nested pre-test attainment improve the power of analyses even if the covariate is not statistically significant (e.g., Raudenbush et al., 2011) and because we explicitly sought to test the Simplicity x PA interaction in predicting growth in attainment even after controls for pre-test reading ability. Equations 1, 2, and 3 describe this final model at the student, classroom, and school levels, for student i in classroom j in school k, respectively, with equation 4 describing the mixed model.

(1) Equation for Student Level 1 Model:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(phonological\ awareness)_{ijk} + e_{ijk}$$

(2) Equations for Classroom Level 2 Model:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(pretest\ attainment)_{jk} + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k,}$$

(3) Equations for School and Province Level 3 Model:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(province)_k + \gamma 002(intervention)k + \mu_{00k}$$
$$\beta_{01k} = \gamma_{010}$$
$$\beta_{10k} = \gamma_{100} + \gamma_{101}(intervention)_k$$

(4) Mixed Model:

$$Y_{ijk} = \gamma_{000} + \gamma_{001}(province)_k + \gamma_{002}(\text{intervention})_k + \gamma_{010}(\text{pretest attainment})_{\text{jk}}$$
$$+ \gamma_{100}(phonological\ awareness)_{ijk} + \gamma_{101}(phonological\ awareness)_{ijk} * (\text{intervention})_k$$
$$+ r_{0jk} + \mu_{00k} + e_{ijk}$$

In all analyses, predictor variables were grand mean-centered except for the bivariate condition and province terms. These exceptions were made to facilitate interpretation of the intercept and the binary cases. Identical 3-level models were run for all variables with mid-test WRAT Word Reading classroom mean as the covariate.

The means and standard deviations for the two conditions are presented in Table 2. Where a test yielded standard scores they are also reported. Inspection of these data suggests few advantages at post-test for the Simplicity program over the Control program at post-tests on WRAT Word Reading, Woodcock-Johnson Word Attack and Spelling, and GRADE Sentence Comprehension measures compared to the mid-test immediately prior to intervention.

Results of the HLM analyses are reported in Table 3 where researcher created dummy-coded Simplicity is compared against the zero-coded Control condition. Dummy codes were also created for the effect of province, where Alberta served as the zero reference category.

The Simplicity intervention involved teaching complex GPC correspondences so that children could read the many words containing these correspondences. Reflecting content, our primary outcome measures were WRAT Word Reading and Woodcock-Johnson Word Attack, so unadjusted alpha was applied for these two measures. All other measures were secondary outcomes and alpha adjustments were made for the total number of secondary outcomes ($\alpha = .05/3 = .017$). Research predictions were specific and directional so planned comparisons were undertaken.

The results revealed no significant main effects of Intervention. Effect sizes expressed as post-test mean differences over pooled post-test standard deviations for WRAT reading, Woodcock–Johnson Word Attack, Spelling, GRADE vocabulary and sentence comprehension were 0.08, −0.27, 0.02, 0.02, and 0.08 respectively. There was, however, a significant Intervention x Phonological Awareness interaction effect for WRAT Word Reading ($p < .01$), Woodcock-Johnson Word Attack ($p < .05$), and Sentence Comprehension ($p < .05$) at post-test. In each case, this favored Simplicity over Control for the children with stronger phonemic awareness ability at pre-test (see Figures 1–3). Other effects did not reach significance. Finally, we re-ran all analyses excluding the two schools added after randomization and the results remained the same.

Table 2. Means and standard deviations for pre- and post post-test literacy measures by intervention group.

	Simplicity Control		Simplicity Control	
Measure	Pre-test	Post-test	Pre-test	Post-test
WRAT[a]	85.68 (9.75)	89.36 (10.55)	86.39 (9.37)	90.26 (11.20)
Word Attack[a]	93.87 (8.32)	94.13 (9.64)	95.13 (10.06)	96.58 (8.49)
Spell[a]		91.18 (9.37)		91.35 (10.23)
Vocab[a]		82.22 (12.14)		82.53 (12.58)
Sentence Comp[b]		2.66 (1.68)		2.54 (1.48)

Note: Values are represented by (a) standard scores, (b) stanines.
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 GRADE Reading Test Vocabulary sub-test
Sentence comp GRADE Reading Test Sentence Comprehension sub-test

Table 3. HLM results for the effect of intervention condition by phonological ability on post-test attainment.

	Child-Level		Classroom-Level		School/Province-Level	
Variable	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
A) WRAT word reading Post = Dependent Variable						
Intercept	29.22	(0.75)***				
Phonological Ability	0.32	(0.15)*				
Intervention	1.12	(1.21)				
Intervention x Phonological Ability	0.37	(0.16)**				
WRAT word reading pre-test			0.58	(0.19)**		
Province					−3.65	(1.13)**
Variance components						
Child	17.42	(4.17)				
School					3.74	(1.93)
B) Woodcock-Johnson Word Attack Post = Dependent Variable						
Intercept	11.79	(0.72)***				
Phonological Ability	0.26	(0.11)*				
Intervention	0.26	(1.01)				
Intervention x Phonological Ability	0.24	(0.12)*				
WRAT word reading pre-test			0.45	(0.19)*		
Province					−1.60	(1.03)
Variance components						
Child	15.69	(3.96)				
School					2.36	(1.54)
C) Woodcock-Johnson Spelling = Dependent Variable						
Intercept	23.47	(0.72)***				
Phonological Ability	0.21	(0.11)*				
Intervention	0.44	(0.49)				
Intervention x Phonological Ability	0.03	(0.11)				
WRAT word reading pre-test			0.46	(0.10)***		
Province					−1.79	(0.96)
Variance components						
Child	7.81	(2.80)				
School					0.09	−0.29
D) GRADE Vocabulary = Dependent Variable						
Intercept	48.15	(1.80)***				
Phonological Ability	0.88	(0.28)**				
Intervention	0.72	(1.20)				
Intervention x Phonological Ability	0.35	(0.26)				
WRAT word reading pre-test			1.04	(0.24)***		
Province					−10.12	(2.25)***
Variance components						
Child	88.99	−9.21				
School					0.02	−0.12
E) Post-test GRADE Sentence Comp = Dependent Variable						
Intercept	10.49	(0.72)***				
Phonological Ability	0.2	(0.09)*				
Intervention	0.46	(0.64)				
Intervention x Phonological Ability	0.19	(0.08)*				
WRAT word reading pre-test			0.41	(0.20)*		
Province					−2.87	(0.71)***
Variance Components						
Child	18.26	−4.27				
School					0.01	−0.19

Note. HLM = hierarchical linear modeling; WRAT = Wide Range Achievement Test III, Reading subtest; GRADE = Group Reading Assessment and Diagnostic Evaluation.*** $p < .001$, ** $p < .01$, * $p < .05$

Discussion

The present study explored the hypothesis that an intervention teaching GPCs derived from the application of the Simplicity Principle would impact performance on standardized reading and spelling tasks over a matched intervention where children were exposed to the same GPCs but only letter names were used to label the GPCs. From previous evidence (Hulme et al., 2012), we

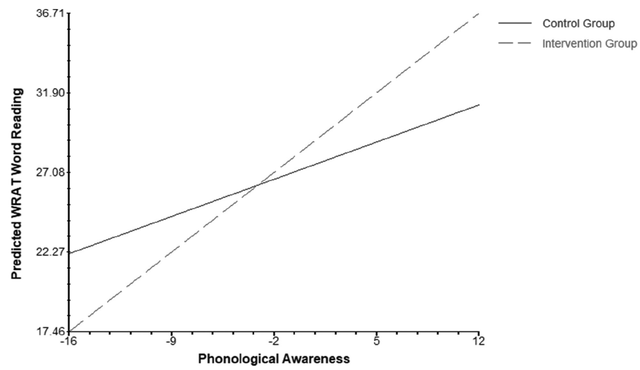

Figure 1. The effects of phonological awarness on WEAT word reading by intervervention.

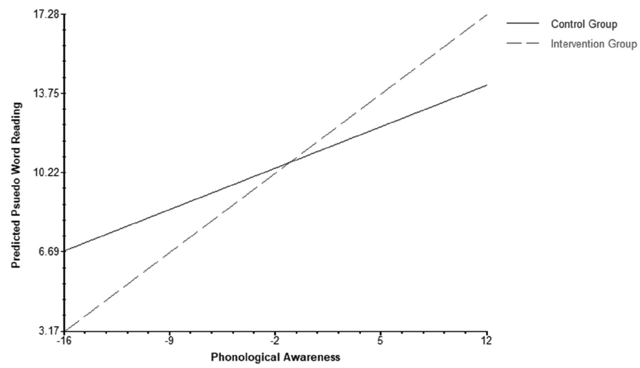

Figure 2. The effects of phonological awarness on WJ Pseudo word reading by intervervention.

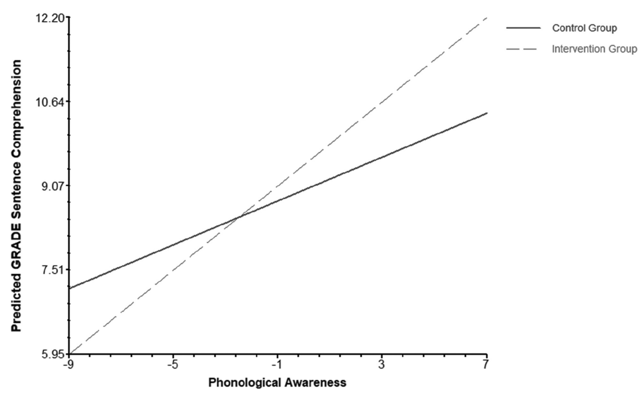

Figure 3. The effects of phonological awarness on sentence comprehension by intervervention.

predicted that the effects of intervention would be strongest where children had stronger phonemic awareness skills. We used pre-test phonemic awareness as a moderator of outcomes and explicitly predicted an interaction between pre-test phonemic awareness and exposure to GPCs on subsequent reading development. We specifically predicted these interaction effects would still hold even after keeping pre-test reading ability constant. This hypothesis was supported by some of the present data. Significant effects that were highly specific to the more phonologically able children were evident on our two primary literacy outcomes and one secondary outcome, sentence comprehension, in comparison to a taught control intervention identical in content except for the explicit attention drawn to the pronunciation of GPCs.

Our study replicates and substantially extends the only other intervention study exploring GPC teaching using the Simplicity Principle (Chen & Savage, 2014). The current findings place research in this domain on a substantially stronger empirical footing because the present study assessed performance in a matched control dual-site trial drawing children from 44 classrooms in 21 schools. Chen and Savage (2014) drew children from 2 classrooms in one school. This larger study was also able to formally test for an interaction between the complex GPCs taught and underlying phonemic awareness skills. The interpretation of results is strengthened by the inclusion of a taught control group who received the same small-group intervention as the main intervention but where only letter names were used to identify the GPC spelling units.

While the idea that reading is co-determined by GPC knowledge and phonological awareness is well established in a general sense (e.g., Byrne, 1999; Hjetland et al., 2017; Hulme et al., 2012; National Reading Panel, 2000), it was important to establish the value and the possible limitations on the application of the Simplicity Principle not established in prior smaller scale work. Results clearly confirm that phonemic awareness moderates the impact of teaching using a Simplicity Principle. This general pattern of moderation of effects of phonics intervention by phonemic awareness in at-risk readers replicates results reported by other teams (see e.g., Hulme et al., 2012; Schaaars et al., 2017). The combination of phonic intervention plus mediation analysis undertaken by Hulme et al. (2012) described earlier has been interpreted as suggesting causal links between both GPCs and phonemic awareness on one hand and reading on the other. Data presented here are consistent with the view that phonemic awareness and GPCs together drive growth in reading following a principled intervention teaching complex GPCs. How then does phonemic awareness facilitate reading development? Hulme et al. (2012) argued that: "to master the alphabetic principle, children need phonemically structured representations of speech as well as letter sound knowledge" (p. 576). We suggest that such representations of speech allow children to connect otherwise abstract declarative knowledge of GPCs to strong letter representations. Such a proposal is also consistent with Schaaars et al.'s (2017) proposal that phonemic awareness operates to make GPC-based decoding and the establishment of stable lexical representations, more efficient.

Finally, while this study does not directly speak to Suggate's (2010) notion about diminishing returns of phonics after Grade 1 because children here were *not* taught phonics skills of how to use GPCs to read and spell words, it does suggest that teaching GPCs, a common component of synthetic phonics programs, is useful in Grade 2, albeit with complex GPCs and for at-risk children with stronger blending skills.

Our approach here has been described as optimizing the learner's statistical learning environment through increasing the chance of item-level learning (Steacy, Elleman, & Compton, 2017). It should be noted, however, that the optimality of the Simplicity–derived GPCs over other candidate GPC schemes remains to be established empirically through interventions with children. Meaningful conceptual links can also be made from our approach of making the statistical properties of print-sound relations apparent, to connectionist modeling of orthography-to-phonology rules and to the statistical learning processes they embody (e.g., Coltheart et al., 2001; Grainger & Ziegler, 2011; Plaut, McClelland, Seidenberg, & Patterson, 1996). Developmentally, if fluent word reading reflects the automation of underpinning GPCs (Ehri, 2014), then using a frequency-based approach for explicit teaching of these GPCs is likely to be helpful in moving at-risk readers with sufficient phonemic awareness to a consolidated alphabetic stage of word reading. Here, GPCs were taught regularly, in conjunction with the reading and spelling of words, and with shared passage reading where taught GPCs were frequently represented, intimately connecting the two processes (Savage et al., 2018). Such approaches potentially bootstrap wider development and are potentially "generative" of the rich orthographic knowledge beyond specific GPCs that is associated with typical development (Compton et al., 2014; Steacy et al., 2018). The interaction we report between phonemic awareness and the GPC-taught intervention producing significant growth in word reading, pseudoword decoding, and sentence comprehension only for the more phonologically able suggests that attempts to create an optimal learning environment based on GPC frequency alone, while sufficient for the phonemically aware, will not in themselves suffice for children with weaker phonemic awareness. Pedagogical support for the development of both phonemic awareness and complex GPC knowledge of the weakest readers in Grade 2 is suggested.

Some limitations of our study are worth noting. First, results speak only to the utility of teaching complex GPCs to at-risk readers in Grade 2 compared to an intervention that showed children letter names for target spellings. The Simplicity GPC intervention was by no means a "magic bullet" as while all coefficients for our condition by phonological ability interaction consistently showed advantages favoring the Simplicity intervention, effects were not significant for any secondary outcome measures except sentence comprehension. Post-test stanine scores on the sentence comprehension task, while significantly higher in the Simplicity condition than for the Control condition, nevertheless remained more than 1.5 standards deviations below average. Many children did not improve in word reading and spelling ability in this study. Improved Simplicity Principle-inspired interventions, especially for those with poor phonological awareness abilities, are suggested. We did not teach variable vowels, or other "strategies" for reading words with complex or inconsistent GPCs that are known to be helpful (Compton et al., 2014; Savage et al., 2018; Steacy et al., 2018; Steacy, Elleman, Lovett, & Compton, 2016). We did not explicitly teach children how to use taught GPCs to read or spell. It may well be that phonics instruction, for example teaching children how to blend Simplicity–derived GPCs to read and spell words, produces larger effect sizes than those reported here. Finally our analysis model was based on the study as a well-matched quasi-experiment with school-level randomization, and would benefit from replication in an RCT.

In sum, this is the first larger scale study to take a principled approach to developing and delivering complex GPC progressions within an intervention. GPC progressions were based on statistical corpus analysis and were then used to teach complex GPCs to at-risk readers in Grade 2. Results suggest that teaching complex GPCs is of value over interventions that do not teach these GPCs and instead teaches letter names. Effects of intervention are, as predicted, moderated by phonological awareness skills. These results provide some preliminary support for using computational approaches to modeling potentially optimal GPC content underpinning advanced phonics programs based upon the content of authentic children's books. The most direct educational implication is that some at-risk children in Grade 2 with stronger phonemic awareness may benefit from being taught complex GPCs derived from the Simplicity Principle, in interventions delivered in the manner outlined above. This work also sets the stage for exploring optimal interventions for children who were non-responders or where low phonemic awareness itself does not respond to wider treatment, and where the approach might fit within contemporary RtI frameworks more generally (Fuchs, Fuchs, & Compton, 2012). Future programmatic work should explore the optimality of this Simplicity-derived GPC content over other candidate GPC schemes, and whether combining phonological awareness training with Simplicity-derived GPC progression and incorporating phonics is efficacious for children with low phonological awareness, and in comparison with alternative approaches to intervention. Future research might also usefully explore the effects of greater content repetition and "intensity" of delivery, impact of such interventions on diagnosed dyslexics, and confirmation of sustained effects of intervention in delayed post-tests. The first phase of this work is currently underway.

Acknowledgments

This work was approved and executed after approval from the McGill University Research Ethics Board Office (REB # 458-0612– Responding to intervention: schools where all children are taught to read), and in full accordance with McGill University policy and the Canadian Tri-Council Policy Statement *Ethical Conduct for Research Involving*

Humans. All participants (or their legal guardians) gave their informed consent and children gave their assent prior to their inclusion in the study. There is no conflict of interest for any of the authors in this work.

Funding

This work was supported by the Max Bell Foundation [none given by Max Bell]; Social Sciences and Humanities Research Council of Canada [435-2012-0164].

Q6

ORCID

Robert Savage http://orcid.org/0000-0002-5645-6002
Rauno Parrila http://orcid.org/0000-0003-4250-8980

References

Al Otaiba, S., Kim, Y.-S., Wanzek, J., Petscher, Y., & Wagner, R. (2014). Long-term effects of first grade multitier intervention. *Journal of Research in Educational Effectiveness*, *7*, 250–267. doi:10.1080/19345747.2014.906692

Bowers, J. S., & Bowers, P. N. (2018). Progress in reading instruction requires a better understanding of the english spelling system. *Current Directions in Psychological Science*, *27*, 407–412. doi:10.1177/0963721418773749

Byrne, B. (1999). *The foundations of literacy*. London: Routledge.

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

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256. doi:10.1037/0033-295X.108.1.204

Compton, D. L., Miller, A. C., Elleman, A. M., & Steacy, L. M. (2014). Have we forsaken reading theory in the name of 'quick fix' interventions for children with reading disability? *Scientific Studies of Reading*, *18*, 55–73. doi:10.1080/10888438.2013.836200

Daniels, P. T., & Share, D. L. (2018). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading*, *22*, 101–116. doi:10.1080/10888438.2017.1379082

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* (*4th* ed.). Pearson: San Antonio Texas.

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

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*, *78*(263), 279. doi:10.1177/001440291207800301

Gontijo, P., Gontijo, I., & Shillcock, R. (2003). Grapheme–Phoneme probabilities in British English. *Behavior Research Methods*, *35*, 136–157. doi:10.3758/bf03195506

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*, 54. doi:10.3389/fpsyg.2011.00054

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. doi:10.2307/1131364

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

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

Henbest, V. S., & Apel, K. (2017). Effective word reading instruction: What does the evidence tell us. *Communication Disorders Quarterly*, *39*, 303–311. doi:10.1177/1525740116685183

Hjetland, H. N., Brinchmann, E. I., Scherer, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: A systematic review. *Campbell Systematic Reviews*, *14*. doi:10.4073/csr.2017.14

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

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F., & Snowling, M. J. (2012). The causal role of phoneme awareness and leter-sound knowldege in learning to read.: Combining intervention studies with mediation analyses. *Psychological Science*, *23*, 572–577. doi:10.1177/0956797611435921

Jimerson, S. R., Burns, M. K., & Van Der Heyden, A. M. (Eds.). (2016). *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.). New York, NY: Springer.

Machin, S., & McNally, S. (2008). The literacy hour. *Journal of Public Economics*, *92*, 1441–1462. doi:10.1016/j.jpubeco.2007.11.008

Mellard, D. F., & Johnson, E. (2008). *RTI: A practitioner's guide to implementing response to intervention*. Thousand Oaks, CA: Corwin Press.

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

Reynolds, M., Wheldall, K., & Madelaine, A. (2011). Early identification of young struggling readers: Preliminary benchmarks for intervention for students in years one and two in schools in New South Wales. *Australian Journal of Learning Difficulties*, *16*, 127–143. doi:10.1080/19404158.2011.586711

Savage, R., Georgiou, G., Parrila, R., & Maiorino, K. (2018). Preventative reading interventions teaching direct mapping of graphemes in texts and set-for-variability aid at-risk learners. *Scientific Studies of Reading*, *22*, 225–247. doi:10.1080/10888438.2018.1427753

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

Schaaars, M. M. H., Segers, E., & Verhoeven, L. (2017). Word decoding development during phonics instruction in children at risk for dyslexia. *Dyslexia*, *23*, 141–160. doi:10.1002/dys.1556

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin & Review.*, *22*, 1614–1629. doi:10.3758/s13423-015-0835-2

Schneider, W., Roth, E., & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology*, *92*, 284–295. doi:10.1037/0022-0663.92.2.284

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568. doi:10.1037/0033-295X.96.4.523

Smith, M. W., Brady, J. P., & Clark-Chiarelli, N. (2008). *Early language and literacy classroom observation K-3 tool* (Research ed.). Baltimore: Paul H. Brookes.

Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology*, *81*, 1–23. doi:10.1111/j.2044-8279.2010.02014.x

Steacy, L. M., Compton, D., Petscher, Y., Elliott, J. D., Smith, K., Rueckl, J. G., … Pugh, K. R. (2018). Development and prediction of context-dependent vowel pronunciation in elementary readers. *Scientific Studies of Reading*. doi:10.1080/10888438.2018.1466303

Steacy, L. M., Elleman, A. M., & Compton, D. (2017). Opening the 'black box' of learning to read: Inductive mechanisms supporting word acquisition development with a focus on children who struggle to read. In K. Cain, D. Compton, & R. Parrila (Eds.), *Theories of reading development* (pp. 99–124). Amsterdam: John Benjamins.

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., & Stainthorp, R. (2015). *Reading development and teaching*. Los Angeles: Sage.

Suggate, S. P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology*, *46*, 1556–1579. doi:10.1037/a0020612

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

Vousden, J., Ellefson, M., Chater, N., & Solity, J. (2010). *Comparing the content of UK reading programs using the simplicity principle*. Poster presented at Seventeenth Annual Conference of the Society for the Scientific Study of Reading, Berlin, Germany, July.

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

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing–2nd ed. (CTOPP-2)*. Austin, TX: Pro-Ed.
Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement test 4*. Lutz, FL: PAR.
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, IL: Riverside Publishing.