JARL
JOURNAL OF APPLIED RESEARCH ON LEARNING
Vol. 1, # 1, Article 2, April 2007

Potential risks to reading posed by high-dose phonics

By Linda M. Phillips, Stephen P. Norris & Dorothy J. Steffler

The meaningful applied phonics (M.A.P.) program is a prescriptive, explicit, teacher-directed approach, designed to introduce the alphabetic principle by teaching the 70 graphemes that make up the 26 letters, and the strategies used to segment and blend words into syllables. Seven schools were selected by a western Canadian school board to implement the program and form the treatment group. The treatment group was taught the M.A.P. program and was followed in a longitudinal study from Grades 1 to 3, and their performance was compared to that of a matched control group who were taught a more balanced program. The seven schools maintained their relative ranking, although their performance declined when compared to the normative samples used for the criterion measures. No systematic effect of gender on achievement was found. The children in the control group outperformed the children in the treatment (M.A.P.) group.

Effective reading instruction is a topic of abiding interest and debate. Many children who should be capable of reading well given adequate instruction cannot do so, which suggests that the instructional methods available to them are not appropriate (Pressley, 1998). Furthermore, the media (Fine, 2001) and some research literature (Francis, 2005; Jha & Kelleher, 2006; PISA, 2005) claim that girls outperform boys and that boys therefore need alternative programs and instruction. In addition to teaching methods and gender, there are a number of other factors that affect instruction, such as SES, ESL, and cultural differences.

There is no shortage of programs and methods dedicated to improving early reading instruction. However, the specific strengths and weaknesses of many of these programs and methods are largely unmeasured. Indeed, very few studies have explored the efficacy of early literacy programs and methods within highly defensible parameters. The subject of this three-year longitudinal study of students in Grades 1 to 3 is a prescriptive, linear, and intensive phonics-based reading program, *Literacy M.A.P. Meaningful applied phonics: Explicit phonics through direct instruction (*Hunter & Robinson, 2002). We focus on three questions from a larger set of questions addressed by Phillips, Norris, & Steffler (2005):

1) What were the treatment children's reading, writing, and spelling achievements over the three years?
2) Was gender related to achievement in the treatment group?
3) Was there an effect of treatment on achievement when the treatment group was compared to a control group?

A large western Canadian urban school board selected seven schools to test the M.A.P. program. The funders, members of the school board and of the Society for the Advancement of Excellence in Education, set the research questions.

Background

Research to date does not provide definitive answers regarding the amount of emphasis that should be placed on phonics or the best way to teach phonics (Cunningham, 2003; National Institute of Child Health and Human Development [NICHHD], 2000; Stahl, Duffy-Hester, & Stahl, 1998). Debates on these issues are sometimes fuelled by confusion about the most widely used terms: "phonological awareness," "phonemic awareness," and "phonics."

Phonological Awareness/Phonemic Awareness

"Phonological awareness" and "phonemic awareness" both refer to an awareness of the sounds (phonemes) that make up spoken words. Phonological awareness, the broader term, is awareness of the constituent sounds of words (Harris & Hodges, 1995). Phonemic awareness focuses on the ability to differentiate and manipulate individual phonemes represented by letters and letter clusters within words (Williams, 1995). Following the example of Blachman (2000), Gillon (2004), and Snow, Burns, and Griffin (1998), we use the more inclusive term "phonological awareness" (PA), because phonemic awareness may be considered a subset of this skill.

Numerous studies of PA conducted over the past three decades indicate that it has a strong positive and powerful effect on subsequent reading achievement (Bradley & Bryant, 1983; Byrne & Fielding-Barnsley, 1995; Lundberg, Frost, & Petersen, 1988; Lundberg, Olofsson, & Wall, 1980; Perfetti *et al.*, 1987; Stanovich, 1986; Torgeson, Wagner, & Rashotte, 1994; Vellutino & Scanlon, 1987). Not only is there a clear indication that PA is related to reading success, but it also appears that learning to read helps develop children's PA (Blachman, 2000; Bus & van IJzendoorn, 1999; Byrne & Fielding-Barnsley, 1989; Ehri *et al.*, 2001; Perfetti *et al.*, 1987; Wagner, Torgesen, & Rashotte, 1994).

Based on research in the field, Gillon (2004) and Snow, Burns, and Griffin (1998) outlined an instructional PA sequence. Snow, Burns, and Griffin report that until children understand the concept of syllables, "their only option for learning to read or spell words is by rote memorization" (p. 54). Furthermore, they maintain that "until children have a basic awareness of the phonemic structure of language, asking them for the first sound in the word *boy* or expecting them to understand that *cap* has three sounds while *camp* has four is to no avail" (p. 54).

Research studies on PA and the meta-analyses of these studies have implications for instruction. Bus and van IJzendoorn (1999) conclude that preschoolers tend to profit most from PA training. The National Reading Panel (NRP) of experts named by the Director of the National Institute of Child Health and Human Development (NICHHD, 2000) argue that the reason kindergartners and preschoolers gain the most from such training is that these groups start out with the least phonological awareness. The NRP included a length-of-instruction variable in its meta-analysis and found that 5 to 18 hours of PA instruction were more effective than longer or shorter periods. Ehri *et al.* (2001) stated, "These findings suggest that phonological awareness instruction does not need to be lengthy to exert its strongest effect on reading and spelling" (p. 269). Bus and van IJzendoorn (1999) stress that although "phonological awareness

affects learning-to-read processes in a positive and substantial way" (p. 405), PA is not sufficient for learning to read. They estimate that PA explains approximately 12% of the variance in reading skills. Ehri *et al.* (2001) calculate the overall variance in reading outcomes explained by PA instruction to be 6.5%, rising to 10% when letters were added and to 28% for preschoolers. Thus, while PA has been shown to be important in reading words, it is but one component among a complex array of literacy experiences that help children learn to read (Cunningham 2001; Pressley *et al.*, 2001, Snow, Burns, & Griffin, 1998).

Phonics

When letter-sound correspondence instruction is added to phonological awareness training, one enters into the domain of phonics. As described by Strickland (1998), phonics refers to instruction intended to develop an understanding of the alphabetic principle (i.e., that letters represent sounds), as well as knowledge of the sounds represented by those letters or letter combinations.

There are a number of practices that have received recognition in the field of reading research. An early introduction to letter-sound correspondences helps children learn to read words. The NRP meta-analysis determined that phonics instruction is most effective when introduced in kindergarten and Grade 1. In a study of four classrooms, Juel and Minden-Cupp (2000) found "phonics first and fast" helped children with low reading skills make substantial gains in successful word reading over the course of Grade 1. This strategy for introducing phonics is "first" because these children spent more than one-third of their language arts time in the fall engaged in phonics activities and "fast" because phonics instructional time fell appreciably by February when reading instruction began to focus more on vocabulary and text meaning. This phonics first-and-fast practice corresponds with that recommended by Stahl (1992), who observed that once students regularly begin to use sound-letter correspondences and orthographic patterns in their reading, it is time for them to spend more time on reading and writing and less time on learning basic phonics. And despite Juel's and Minden-Cupp's finding that intensive phonics instruction helped students with low reading skills make substantial gains, Juel *et al.* (2003) advise that preoccupation with teaching decoding skills may have a negative impact on students' overall reading ability: decoding, although crucial, is not sufficient.

Systematic instruction in phonemic awareness is widely advocated (Morrow & Asbury, 2003; NICHHD, 2000; Pressley, 2002; Strickland, 1998). At the present time, because research has yet to indicate any one highly effective system for phonics instruction, "organized and comprehensive" may best describe the type of phonics instruction that the research base supports. Although most early reading programs pay at least some attention to phonemic awareness, the M.A.P. program, under study here, has a very high dosage.

A Case for Balanced Programs in Reading

After a detailed presentation of the reasons why students need to develop "a deep and thorough knowledge of letters, spelling patterns and words, and of the phonological translation of all three" (p. 416), Adams (1990) concluded that these "must be developed in concert with real reading and real writing and with deliberate reflection on the forms, functions, and meaning of texts" (p. 422). Both the National Reading

Panel (NICHHD, 2000) and the committee of reading experts who produced *Preventing Reading Difficulties in Young Children* (Snow, Burns, & Griffin, 1998) came to a similar conclusion. The NRP cautioned frequently that although they found the development of PA and phonics skills to be a necessary component of reading programs, children need to acquire additional skills if they are to become competent readers and writers. They advise that programs focusing too heavily on the teaching of letter-sound relations and not enough on putting this knowledge to work in actual reading activities are unlikely to be very effective.

Pearson (2001) includes among his fundamental tenets about reading the principle that skills development is an essential feature of reading and writing instruction and that, since reading is the whole point of reading instruction, "a curriculum that postpones real reading for more than an instant does kids a disservice by raising in their minds the possibility that reading may not be the point of reading instruction" (p. 80). This stance puts him in the company of those who advocate a balanced view of reading, a balance he does not consider to be an attempt to balance off aspects of instruction endorsed by different reading factions (such as phonics vs. whole language advocates), but as "assembling an array of skills, strategies, processes, and practices that are sufficiently rich and synergistic to guarantee a full and rich curriculum for all students" (p. 82). In sum, it must be remembered that phonics instruction is a means to an end, that being the ability not just to decode but to comprehend and interpret written text.

Method

Design and Sample

The study compared a treatment group initially comprising 272 students from seven schools to a control group identified by a large, urban, western Canadian school board. The control group was given no special treatment, and took the program in use in the schools at the time of the study. The treatment group was tested during the last week of September and the last week of May in Grades 1, 2, and 3, so that their performance could be tracked and compared against the sample norms used in the design of the tests. In addition, achievement data collected at the end of Grades 1, 2, and 3 on both the treatment and control groups were obtained from the school board and provincial testing programs. These results were used to test the effect of treatment (see untreated, non-equivalent control group design with pretests and posttests in Cook and Campbell, 1979, pp. 103-118).

The seven schools using the M.A.P. program were identified by the school board to be at risk for reading failure. Six of the schools were urban, and one, rural. The six urban schools represented a range of multicultural diversity (from a low level of 5%-10% for Schools 2 and 3; a moderate level of 35%-40% for Schools 4 and 5; to a high level of 50%-60% for Schools 6 and 7). School 1 served an Indigenous population. Mobility rates per year for the years 1999 to 2004 varied from a lowest level of about 10% for Schools 3 and 7, about 15% for Schools 2 and 5, and over 20% for Schools 4 and 6. No data were available for School 1.

The control group was selected by the school board to match only the students for whom complete school board and provincial data were available from Grades 1 to 3 inclusive (117 students out of 272 who began the study). Control group students were selected to match the treatment students on gender, school size, SES, ESL, and cultural diversity. The control group data were not made available by the school board until the spring following final data collection on the treatment group. Selecting the control group in this manner ensured that the internal validity of the study was not compromised by differential attrition from the treatment and control groups.

Assessment Instruments

The following three measures of students' literacy development were used only with the treatment group children.

Group Reading Assessment and Diagnostic Evaluation (GRA⁺DE).

Reading performance was measured in the fall (Form A) and in the spring (Form B) of each year using the Group Reading Assessment and Diagnostic Evaluation (GRA⁺DE) (Williams, 2001). The GRA⁺DE is a group-administered reading assessment tool and includes ten subtests that measure various components of early reading and pre-reading skills. All subtests are multiple-choice format with four options.

Children in Grade 1 were tested using the Level K GRA⁺DE, as well as selected subtests from Level 1. Level K is designed for early first grade and transitional first-grade classrooms (Williams, 2001). Children were tested using the following eight subtests of Level K (sound matching and rhyming, same and different words, print awareness, letter recognition, phoneme-grapheme correspondence, listening comprehension, and word reading), which are described below.

The sound matching and rhyming subtests measure phonological awareness of beginnings and endings of common nouns that are represented with pictures. Children are asked to mark the picture that starts or ends with the same sound as a given word. For example: "Next to the train are four pictures. Find the one that begins with the same sound as *what*. Listen. The pictures are *dad, arrow, branch,* and *whale*. Look at the pictures. Mark the one that starts with the same sound as *what*" (Williams, 2001, Teacher's Administration Manual, Level K, Form A, pp. 34-38).

The same and different words subtest is a measure of visual skills and early literacy skills. Children are asked to identify one of four words that is either the same as a target word or different from the other three words. The print awareness subtest is a measure of visual skills and knowledge of print material. Children are asked to identify which one of four items correctly answers a given question. For example, "Find the box that has a *sentence* in it" or "Find the box with a *capital* letter in it" (Williams, 2001, pp. 30-31). The letter recognition subtest measures visual skills and knowledge of print material. Children are required to identify specific letters written in either upper- or lower-case font (Williams, 2001, pp. 48-49). These three subtests combine to form an early literacy skills measure.

The phoneme-grapheme subtest measures visual and auditory early reading skills. Children are asked to identify the printed letter that matches either the beginning or ending sound of a common word. For example: "Mark the letter that makes the sound that you hear at the end of *get*" (Williams, 2001, p. 79). The listening comprehension subtest is a measure of understanding spoken language with no print cues. Children are asked to mark the picture that best fits a spoken sentence. For example, the child might be given the sentence, "The horse is jumping over the fence," and asked to find the appropriate picture (Williams, 2001, p. 51). The word reading subtest measures the child's ability to recognize grade-appropriate sight words and decode simple, regular words. Children are asked to mark the word that the tester reads. For example, "Big. The hat is too big. Mark big" (Williams, 2001, p. 83).

The three subtests used from the Level 1 GR*A*$^{+}$DE were word meaning, sentence comprehension, and listening comprehension. The word meaning subtest measures word decoding and understanding grade-appropriate reading vocabulary. Children are asked to read a word silently and identify the picture that represents what the word means. The sentence comprehension subtest measures the child's ability to comprehend an entire sentence or complete thought. Children silently read a given sentence with a missing word and choose the appropriate word to fill in the blank. The listening comprehension is a more advanced version of the corresponding kindergarten subtest.

Children in Grade 2 were tested using the following Level 2 GR*A*$^{+}$DE subtests: listening comprehension, word reading, sentence comprehension, word meaning, and passage comprehension. The listening comprehension, word reading, sentence comprehension, and word meaning subtests for Level 2 are similar in format and administration to the Level 1 subtests described above. Passage comprehension measures the child's ability to read and comprehend a short passage. Children read the passage silently and answer multiple-choice questions about its content.

Grade 3 children were tested using the following Level 3 GR*A*$^{+}$DE subtests: listening comprehension; word reading and vocabulary forming a vocabulary composite; and sentence comprehension and passage comprehension forming a comprehension composite. The Level 3 listening comprehension, word reading, sentence comprehension, and passage comprehension subtests are similar in format and administration to their Level 1 and 2 counterparts. The vocabulary subtest measures the child's knowledge of word meaning. Children are asked to read a phrase silently and identify the meaning of the word printed in bold font.

Test of Written Spelling (TWS)

Spelling performance for all three grades was measured using the Test of Written Spelling, Fourth Edition (TWS-4) (Larsen, Hammill, & Moats, 1999). Form A was used for fall testing and Form B for spring testing. The TWS-4 is a dictated words group-administered spelling test. The test words are dictated to the children in the typical word-sentence-word format. Test items include both phonologically predictable and unpredictable spellings, for example *pile, knife*.

Oral and Written Language Scales (OWLS)

Writing performance for all three grades was measured using the written expression scale of the Oral and Written Language Scales (OWLS) (Carrow-Woolfolk, 1996). The OWLS measures a child's use of writing conventions, such as letter formation, capitalization, punctuation, ability to use various linguistic forms (such as modifiers, question form, simple and complex sentences), as well as the ability to communicate meaningfully in written form using appropriate content, detail, and coherence. The OWLS is designed for individual or group administration. Test items consist of both structured and open-ended writing tasks. Although students received the same test in each year of testing, the items are developmentally appropriate with suggested start and stop points based on the child's age. The easiest items require students to write their own names and individual letters, copy words and sentences, and write words that match a picture. In Grades 2 and 3, children are required to construct their own sentences, write a complex sentence out of two simple sentences, and ultimately write a short story. The child's performance is scored for letter formation, spelling, capitalization, punctuation, conventional sentence structure, use of modifiers, verb forms, phrases, complex sentences, meaningful content, and detail, in accordance with explicit and specific criteria specified in the administration manual.

The following two measures were available on both the treatment and control group children.

Highest Level of Achievement Tests (HLAT)

During the two-week period from April 26 to May 7, 2004, all students enrolled in Grades 1 to 9 in the school board involved in the study and who received instruction in English wrote the Highest Level of Achievement Tests (HLAT) in reading and writing. Only the results of reading subtest of the HLATS, which consisted of the reading comprehension subtest of the Canadian Test of Basic Skills (Hieronymus, et al., 1998), were released by the school board for use in this study.

Provincial Achievement Tests (PAT)

Provincial Achievement Tests provide performance information about how well students are demonstrating provincial standards in core courses of language arts and mathematics at Grade 3. Only the reading subtest results of the PAT were released by the school board. The purpose of the achievement tests is to determine whether students are learning what they are expected to learn; to inform citizens about students' achievement compared to provincial standards; and to assist schools, school authorities, and the province in monitoring and improving student learning.

The HLAT and the PAT are appropriate dependent measures for comparing the treatment and control groups. Whatever other specific goals might be held for the M.A.P. program, it must have the goal of teaching the more general literacy goals set by the province, which are measured by the HLAT and PAT.

Teacher Survey on Fidelity of Implementation and Teacher Preparation in M.A.P.

A teacher survey was designed, developed, and administered by personnel from the school board to examine the implementation of M.A.P. program elements and program training. Principals at each of the schools were also asked about their perceptions of the M.A.P., but only the teacher survey comments

afforded a systematic and direct indication of actual program implementation. Forty-one teachers were surveyed by a school board consultant. Survey responses from 37 teachers were collected by the school board personnel and sent to us. An independent rater, working for the research team, judged the teachers' qualitative and quantitative responses.

The fidelity of implementation survey included six main categories of items. These included: (a) the number of phonograms (a graphic character or symbol representing a phonetic sound, phoneme, or word) used/taught; the amount of time allotted each day or week for phonograms; the number of spelling words used/taught; the amount of time allotted each day or week for spelling; reproducible charts used/taught; and homework routines, if any. Teachers were invited to provide additional comments. The program training component of the survey questioned teachers' hours of training on the M.A.P. program, hours of coaching or demonstration visits, and years of experience in using M.A.P. (or a related program). Teachers were invited to provide additional comments on this component of the survey as well.

The range of implementation fell short of the full implementation by all teachers and schools that would have provided the most robust test of the M.A.P. program. However, there was no discernible relationship between implementation level and effects of program. The teacher preparation in M.A.P. pointed to a similar picture: it was not possible to tie preparation to success or lack of success of students. Performance in schools was widely divergent despite common levels of teacher preparation in the M.A.P. program.

Procedure

Treatment Group

The treatment group was taught using the *Literacy M.A.P. Meaningful applied phonics: Explicit phonics through direct instruction* (Hunter & Robinson, 2002) program. The same group of students was followed for three years, starting at the outset of Grade 1 in the fall of 2001 and ending in the spring of 2004. The general features of M.A.P. as presented by Hunter and Robinson include: (a) M.A.P. is an explicit, teacher-directed approach; (b) M.A.P. is a logical, sequential program that organizes and paces the lessons, moves from graphemes to spelling words, and to reading and writing activities; (c) M.A.P. segments and blends words into syllables when reading and writing unfamiliar words; (d) M.A.P. teaches the 70 graphemes for the first 26 single letters sounds of the alphabet as well as vowel and consonant digraphs; and (e) M.A.P. focuses on phonics, spelling, and grammar for approximately 30 minutes each day combined with 60 minutes of reading and writing.

There are also more specific features of M.A.P advocated by Hunter and Robinson (2002). The program involves teacher-directed whole-class instruction and a sequential development of skills for approximately 90 minutes each day. This sequence starts with students learning to print correctly the letters of the alphabet following a detailed set of instructions. As they practice writing each letter, they also learn the

sounds of that letter. The M.A.P. authors refer to the instruction students receive when they see and print letters and hear and say sounds as multi-sensory. This multi-sensory component is an important part of the program according to the authors.

After students have learned to print and say the sounds of the alphabet (referred to as "graphemes" in the manual), they follow the same routine to learn 28 multiple-letter "graphemes" (sometimes also referred to as "phonemes" in the manual). For example, students learn such phonemes as –er, –ir, –ur, th (two sounds), and –ou (with its four sounds: ow, ō, oo, and ŭ). Then, children begin using their grapheme knowledge to learn to spell words. The manual suggests graphemes be taught at a rate of four per week in kindergarten or 16 per week in higher grades. As stated in the manual, the process for teaching new spelling words is "very explicit in nature" (p. 19). This process entails listening to the teacher pronounce a word, distinguishing each sound, and writing down the letter(s) associated with that sound. If more than one sound is associated with a grapheme, the grapheme is marked with a number indicating which sound is being used. For example, the sounds of each grapheme are learned in order of frequency, and that order is used for marking the graphemes in a spelling word with an ordinal number placed above each grapheme to indicate its first, second, third, fourth, fifth, or sixth sound. Students are asked what sound they hear first, second, and so on (see p. 18 for more details); they must bear in mind that every syllable must have a vowel, that vowels can have more than one sound, and that only vowels can say their names in a word (see p. 51 for more details). Students must remember there are five spellings of the phoneme –er, five jobs for silent "e," and many other rules. For example, the word "to" with 3 over the "o" (the 3 indicating the third sound of o as in □); 'the' with a line under "th" to indicate that it is a grapheme with two or more letters but one sound and a 2 over the "th" to indicate that the sound is like the "th" in "this" (see p.105).

Students are also given a spelling rule for each word and orally repeat each rule. After students have learned nine spelling words, the teacher incorporates dictated sentences using the mastered words into their phonics lessons. Dictated paragraphs and stories are added when children can spell enough words to proceed to this level of writing. For example, it is recommended that dictated paragraphs not begin until students have been taught the first 54 graphemes, the five jobs for silent "e," and three weeks of dictated sentences. The following is a sample paragraph**:** "My name is _____ _____. I am ___ years old. I go to _____ School. I am in grade one. My teacher is ____ _____"(Hunter & Robinson, 2002, p. 23). The foregoing is consistent with the recommendation that material be presented to students in small sequential steps so as "to provide students with a safe and secure environment from which they can concentrate on the mechanics of writing" (Hunter & Robinson, 2002, p. 25).

The reading process outlined in the M.A.P. model proceeds from the writing process described above. The reading process begins with graphemes taught in isolation. Graphemes are followed by spelling words, dictated sentences, dictated paragraphs and stories, reproducible stories, and, finally, children's literature. The M.A.P. program also incorporates instructions and strategies for teaching fluency, comprehension, and vocabulary. For example, in the section entitled "Comprehension Activities,"

pre-reading discussions, predictions, corrective reading strategies, paraphrasing, mental imagery, and graphic organizers are listed and briefly described. The authors refer to "Best Practices in Promoting Reading Comprehension in Students with Learning Disabilities" (Mastropieri & Scruggs, 1997) for the strategies listed and adapted.

The bulk of the manual contains organizing charts (for kindergarten to Grade 6), instructions on how to print and pronounce graphemes, and spelling word lists. These provide the means "to ensure that teachers and children have the knowledge, skills, and background to succeed on their journey toward literacy" (Hunter & Robinson, 2002, p. 4). Thus, the program manual, like the teaching program, provides explicit directions intended to help teachers to lead children to reading success. The M.A.P. resource is 311 pages and includes program explanations, approximately 200 pages of graphemes, spelling word lists, 115 pages of grapheme sequence lists, and thematic word lists. The reading process model can be conceived as a chain linking graphemes to spelling words, then sight words, dictated sentences, dictated paragraphs and stories, reproducible stories, and, finally, children's literature.

Control Group

The control group was taught according to a balanced literacy approach modeled after Cunningham and Hall's *Four-Blocks Literacy Model* (2001) (this program has two components—Grades 1 to 3 and Grades 4 to 6+; only the former is relevant here). The model is described as incorporating "four different approaches each day to teach children how to become better readers, writers, and spellers" (see www.four-blocks.com, Overview, p.1) and was developed with the needs of a diverse range of students (with various literacy levels, interests, and ways of learning) in mind.

Students are not taught using four-blocks instruction until Grade 1. At that point, literacy instruction is divided into four sections, each requiring approximately the same time: guided reading, self-selected reading, writing, and working with words. Guided reading, which includes learning about story elements and about how to learn from informational texts, "is always focused on comprehension" (Cunningham & Hall, 2001, p. 1, under the Guided Reading block, see www.four-blocks.com). Teachers and students read together big books and the initial reading texts. Teachers introduce other books and formats as the year progresses. Sometimes teachers work with the whole class, but students also read together in small groups, with partners, individually, and with the teacher. During the self-selected reading block, children read what they themselves have chosen. Students are asked to respond to what they have been reading, to share their reading and responses with others, and to conference with the teacher.

The writing block includes mini-lessons on the fundamentals of writing, such as how to get started, revise, and edit their writing. Children are also invited to share their writing and respond to the writing of their peers. The fourth block, working with words, is intended to "ensure that children read, spell, and use high frequency words correctly, and that they learn the patterns necessary for decoding and spelling" (Cunningham & Hall, 2001, p. 1, under the Working with Words block, see www.four-blocks.com). Strategies include making words and a word wall, with suggestions for grouping such activities into a

five-lesson cycle. This approach is intended to help teachers meet mandates for including systematic, sequential phonics in reading instruction. The school board staff indicated that the four blocks are collapsed into three large blocks: working with words, reading, and writing (2001, Carson-Dellosa Publishing Company, available as cited on 11 May 2005 and confirmed 14 March 2007, www.four-blocks.com).

Results and Discussion

What were the treatment children's reading, writing, and spelling achievements over the three years?

The purpose of addressing this question is to provide a sense of how the reading profile of the treatment children changed over the course of the study. This analysis is interesting in and of itself. Even though strictly speaking it is not sound to attribute changes to the M.A.P. program, it is important to see how children's reading developed over three years of instruction using M.A.P. Rather than present all of the data for each of the six points in time over the three years, we present the data at the end of Grade one, then make comparison to the beginning of Grade 1 and to the end of Grade 3 achievement.

Figure 1. Normative Scores for Grade 1 Spring Testing by School

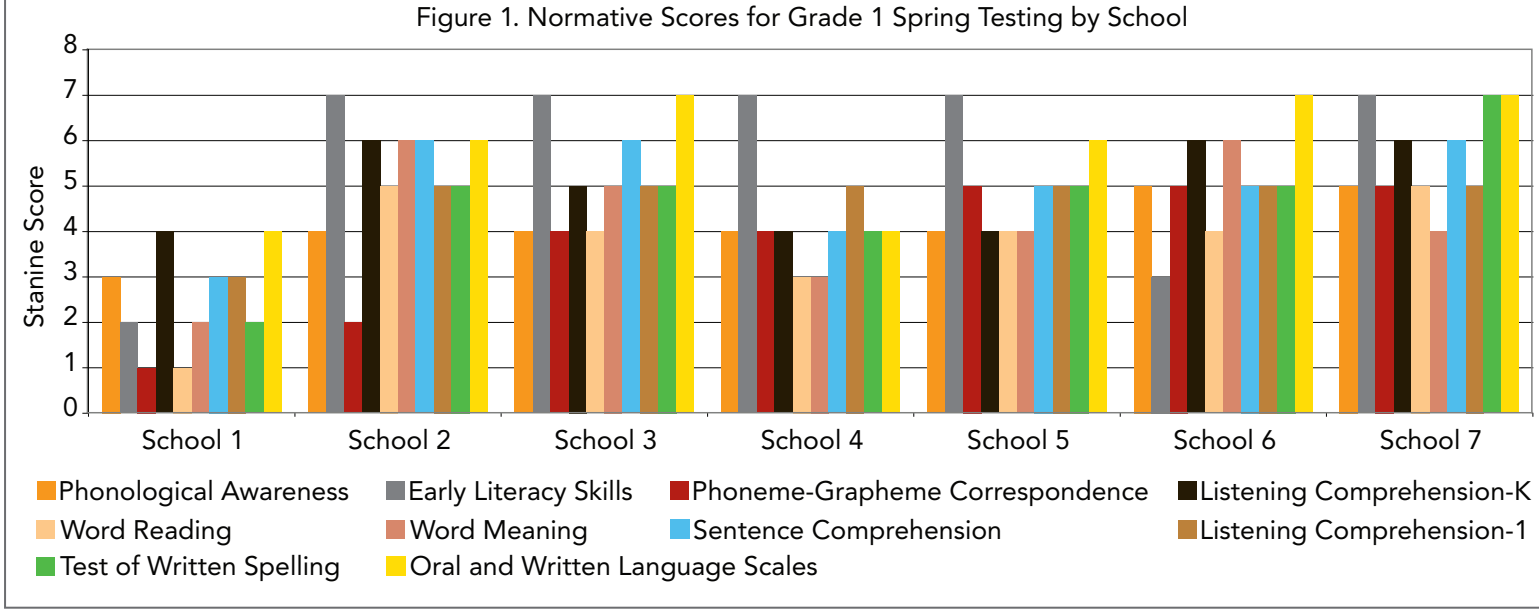


Figure 1 presents the stanine scores for the spring testing by school for ten of the measures (the GRA^{+}DE subtests, the TWS, and the OWLS). We note that School 1 was below the average stanine score of five (representing a percentile rank of 41- 60) on all ten measures, and School 4 was below on nine of the measures. The best performing schools, 2 and 7, were above the middle stanine score on five of the ten measures. Schools 3, 5, and 6 scored above the mean on 3, 2, and 3 of the measures, respectively. This group of schools is thus broadly representative, spanning the range from below average to slightly above average.

For Grade 1, it is interesting to examine the measures on which schools showed improved stanine scores for spring over fall, diminished scores, and scores showing no change. Figure 2 displays the differences between fall and spring performance by school. A positive difference represents improvement compared to the normative sample; a negative difference represents a decrease in standing compared to the normative group. In some cases no differences were found. Schools 3, 4, and 5 showed the greatest improvements on nine, seven, and ten of the measures, respectively, while having diminished performance on zero

measures. Schools 1 and 7, the worst and best performing, respectively, showed the least improvement (on three measures) and the greatest diminution (on four and three measures, respectively). These results suggest that the middle group fared the best. One interpretation of these findings is that the students of School 1 were not ready for the heavy emphasis on phonics offered through the M.A.P. program and that the students of School 7 were sufficiently advanced that the program was of little or no merit.

Figure 2. Stanine Difference between Fall and Spring Testing for Grade 1 by School

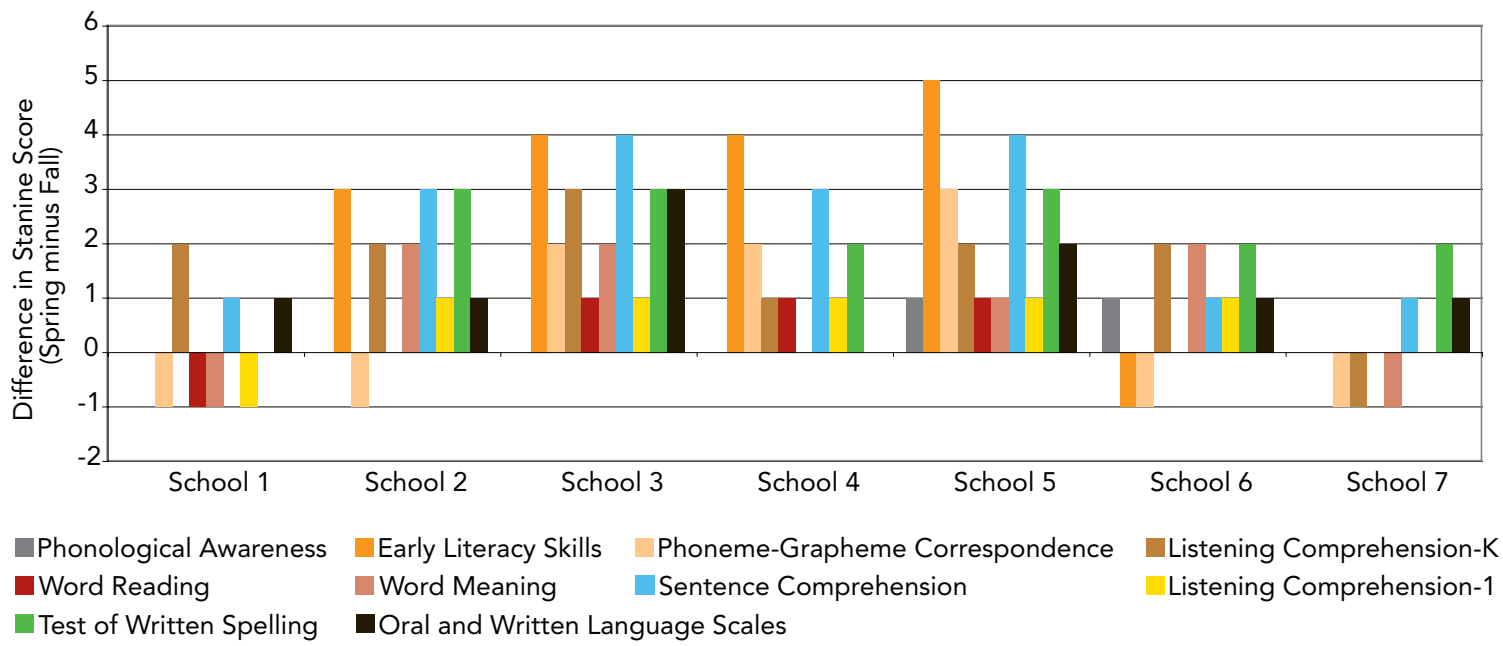
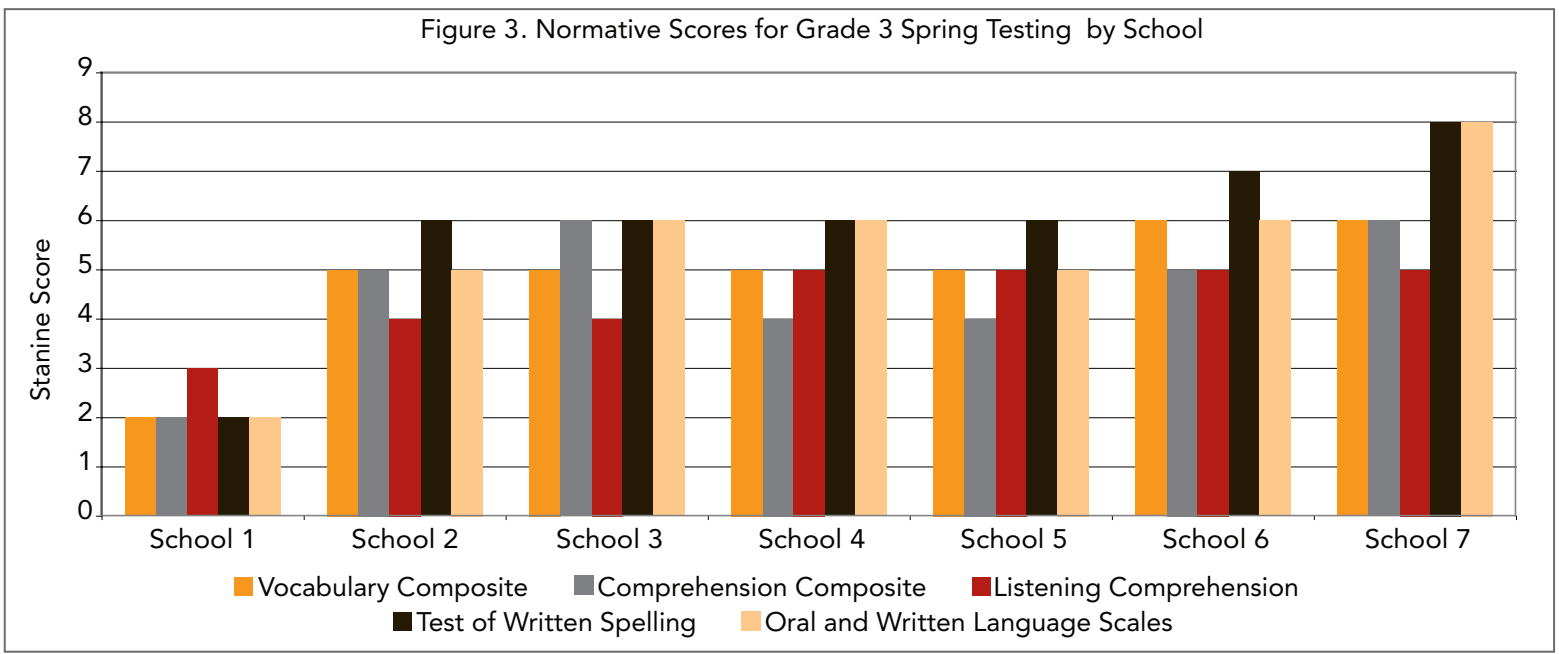


Figure 3 shows how schools rank against the normative standard at the end of Grade 3. School 7 still ranks at the top and School 1 at the bottom. Schools 2 to 6 have become much more alike in performance than they were in the previous two grades. School 4 has caught up and in Grade 3 ranks near the average on the normative scores. However, Schools 2, 3, 5, and 6 have regressed against the normative sample and also rank near the average. In fact, School 1 after three years now looks worse off than at the end of Grade 1 at which time it reached the 4th stanine twice. Overall, we saw less variation in achievement among the

Figure 3. Normative Scores for Grade 3 Spring Testing by School



schools at the end of Grade 3. Not evident from this figure is that it was listening comprehension and oral and written language skills where students' achievement dropped the most, suggesting that knowledge of isolated skills gained through M.A.P. was not translating into genuine literacy performance.

Was gender related to achievement in the treatment group?

Effects of gender were tested at Grades 1, 2, and 3. At Grade 1, in order to test whether the differences between the means of females and males were significant, the ten test scores (the eight diagnostic scores from GRA^{+}DE, not including the subtest scores, the TWS, and the OWLS) were analyzed using multivariate analyses of variance (MANOVA). The multivariate F-test for Gender was not significant for either the fall or the spring testing. Therefore, we concluded that the girls and boys performed equally. Using similar analyses, no significant gender differences were found at Grades 2 and 3.

Was there an effect of treatment on achievement when the treatment group was compared to a control group?

Data for the participants in the M.A.P. program and their counterparts in the control group were available over three years on the Highest Level of Achievement Test (HLAT). In Grade 3, scores were also available on the Provincial Achievement Test (PAT). HLAT scores in Grade 1 were used as a covariate in a multivariate analysis of covariance (MANCOVA). The first between subjects factor was Gender and the second factor was Condition with two levels, Treatment (M.A.P.) and Control (based on Four-Blocks Literacy Model). Observed descriptive statistics for the control and treatment groups are provided in Table 1. The data show the treatment group outscoring the control group at the end of Grade 1 and scoring behind the control group at the end of Grade 3.

It must be noted that, due to restrictions imposed by the school board on data collection and release, we did not have available a measure to serve as a covariate at the beginning of Grade 1. Thus, the covariance analysis equalized the treatment and control groups at the end of Grade 1 after the treatment group had completed one year of the M.A.P. program. This means that any effect of treatment by the end of Grade 1 is masked by the analysis, and only the subsequent effects experienced in Grades 2 and 3 are estimated.

Table 1

Descriptive Statistics by Condition

	Treatment			Control		
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Grade 1 HLAT	117	29.4	9	117	26.7	8.8
Grade 2 HLAT	117	34.2	6.5	117	34.3	6.0
Grade 3 HLAT	117	23.3	5.7	117	23.8	4.8
Grade 3 PAT	117	31.6	6.1	117	32.8	5.0

The MANCOVA results indicated that Grade 1 HLAT was a significant predictor of the three dependent variables taken together—Grade 2 HLAT, Grade 3 HLAT, and Grade 3 PAT, F (3, 227) = 84.31, $p < .05$—as well as for each individually in the corresponding univariate analyses ($p < .05$). Also significant in the multivariate analysis was the condition factor, F (3, 227) = 5.31, $p < .05$. The condition

was also significant for each of the three dependent variables when adjusted for the covariate ($p < .05$). Table 2 displays the predicted means from the fitted model for each of the three dependent variables. In each case, the control participants significantly outperform the treatment participants. Gender and the interaction between gender and condition were not significant. Hence, it can be concluded that males and females performed similarly on the three dependent variables after partialling out the covariate effect and that the participants from the control classrooms performed better than the participants in the treatment classrooms, both collectively and individually, on measures taken in Grades 2 and 3. That is, the M.A.P. students would have done better had they *not* been in the M.A.P. program.

Table 2

Predicted Means for Treatment and Control Groups Using Fitted Model

Variable	Treatment	Control
Grade 2 HLAT	33.0	34.7
Grade 3 HLAT	22.3	24.1
Grade 3 PAT	30.7	33.7

Conclusions and Policy Implications

It is important to remember two limitations of this study when considering conclusions and policy implications:

1) There was no random selection of either schools or students into the treatment and control groups. The control students were matched to the treatment students on the basis of school and socio-cultural characteristics. This limitation affects the external validity of the study and thus inferences beyond the actual students studied warrant caution.

2) The attrition from 272 treatment children at the outset of the study to the 117 at the end (as well as an additional 117 control children) could have influenced results. This decrease does not necessarily mean that students actually left the schools. Rather, it means either that they were not available for some aspect of the testing or that the school board was unable to find an appropriate match. The attrition does not affect the internal validity of the study because the control group was selected after data collection was completed for the treatment group. It does, however, bear upon the external validity of the study for the same reason mentioned in point 1).

We showed that, although the treatment children's reading, writing, and spelling appeared to show some gains in the first year, achievement over the three years deteriorated for children in the lowest- and highest-performing schools (even though the schools maintained their relative ranking as the lowest and highest). Children in the middle performing schools had become more alike in their performance but had also fallen compared to the norming populations. These results raised the questions of whether the children in School 1 were ready for such a high dose of phonics and whether the children in School 7 were too advanced in their reading development for it.

Gender was not found to be related to achievement in the treatment group. This result is consistent with other longitudinal studies we have conducted (Phillips, Norris, & Mason, 1996; Phillips, Norris, Osmond, & Maynard, 2002; Phillips, Hayden, & Norris, 2006), as well as those of other researchers including Biemiller and Siegel (1997), Halpern (2004), and Hyde (2005).

There was an effect of treatment on achievement when compared to a control group.

The control group comparison provides strong evidence that the gains in early literacy achievement experienced by the children in the treatment schools were *less* than would be expected had the children been taught the same program as the control children. The differences were such that the control children, who underperformed the treatment children at the end of Grade 1, outperformed them at the end of both Grades 2 and 3. Clearly, the M.A.P. program (at least during Grades 2 and 3) did not have as beneficial an effect on students' literacy achievement as the programs used in the control schools.

Continued use of the M.A.P. program is indefensible based upon the evidence produced in this study. This recommendation is sound, even if, as was the case with some treatment schools in the study, students were performing at acceptable levels compared to the standards of the school board. This study shows that, however well the treatment students performed in comparison to district averages, they would have performed better with the program used in the control schools. There is an obligation to provide students the program that evidence shows will maximize their performance. Therefore, support for research such as this, which has been provided by the school board, is critically important for the very reason that it has the potential to uncover surprising and perhaps unwanted findings. Unwanted findings are just as important as desired ones in supporting educational policy.

Steps should be taken to allow such policy-relevant research as this to reach a more timely conclusion. In particular, where there is the possibility that children might be disadvantaged through the continued use of some experimental program, comparative data to make this judgment must be available from control groups both at the outset of the research and no later than at the end of the first year. Such steps are not to discourage innovation and experimentation with new programs and methods, which are critically important, but to protect children. Not all innovations and experiments will prove valuable, and for the children's sake that possibility must be kept uppermost in our minds in recommending policy.

Potential innovations also must face strong critical appraisal in light of existing knowledge before they ever make it to trial in the classroom. Some suggested innovations have a known high probability of failure, and for this reason should not be adopted for trial. The body of previous research suggests that any program with as uni-dimensional a focus as the M.A.P. program is likely to be a less effective program for early readers than a multi-dimensional program.

This study will doubtless encourage further vigorous and important debate about reading and reading methodology, especially for diverse and challenging populations. Continued scrutiny of the effects of instructional approaches is a positive outcome of research such as this. It will lead to greater understandings

about successful practices; responsible, data-driven decision-making; and improved literacy programs and outcomes for students.

In conclusion, the M.A.P. program was shown to be detrimental to the children's literacy development in this comparative study. The unfortunate fact is that had the existing research available at the time M.A.P. was developed and implemented been objectively studied, the lesson was clear: phonics is important, but deal with it first and fast and get on with reading. A higher dose of phonics than implied by this maxim is likely to pose a risk to reading. This study supports evidence-based practice in education. Such a practice entails resisting credulity and accepting that even though something seems to make sense, it still might not work. A policy to adopt when evaluating programs is to ask the question: What is the evidence?

About the authors

Linda M. Phillips is Professor and Director of the Canadian Centre for Research on Literacy, Killam Annual Professor at the University of Alberta and was Principal Investigator on this study.

Stephen P. Norris is Professor and Canada Research Chair in Scientific Literacy and the Public Understanding of Science, Department of Educational Policy Studies at the University of Alberta, and was co-investigator on this study.

Dorothy J. Steffler is Associate Professor of Psychology at Concordia University College of Alberta and was the coordinator of the data collection for and co-investigator on this study.

References

Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Biemiller, A., & Siegel, L. (1997). A longitudinal study of the effects of the *Bridge Reading Program* for children at risk for reading failure. Learning Disability Quarterly, 20, 83–92.

Blachman, B.A. (2000). Phonological awareness. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & P. Barr (Eds.), *Handbook of reading research: Vol. III* (pp. 484–502). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Bus, A.G., & van IJzendoorn, M.H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91, 403–414.

Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology*, 81, 313–321.

Byrne, B., & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87*, 488–503.

Carrow-Woolfolk, E. (1996). *Oral and written language scales: Written expression scale*. Circle Pines, MN: American Guidance Service.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company

Cunningham, J.W. (2001). The National Reading Panel Report. *Reading Research Quarterly, 36*, 326–335.

Cunningham, P.M. (2003). What research says about teaching phonics. In L. M. Morrow, L. B. Gambrell, & M. Pressley (Eds.), *Best practices in literacy instruction* (pp. 65–86). New York: Guilford Press.

Cunningham, P.M., & Hall, D. (2001). *Four-Blocks Literacy Model*. Greensboro, NC: Carson-Dellosa Publishing Company.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*, 250–287.

Fine, S. (2001, August 27). Schools told to fix boys' low grades. *Globe and Mail*, p. A1.

Francis, B. (2005). *Reassessing gender and achievement: Questioning contemporary key debates.* New York: Routledge.

Gillon, G. T. (2004). *Phonological awareness: From research to practice.* New York: Guilford Press.

Halpern, D. (2004). A cognitive-process taxonomy for sex differences in cognitive abilities. *Current Directions in Psychological Science, 13,* 135–139.

Harris, T. L., & Hodges, R. E. (Eds.) (1995). *The literacy dictionary: The vocabulary of reading and writing.* Newark, DE: International Reading Association.

Hieronymus, A. N., Hoover, H. D., Lindquist, E. F., Scannell, D. P., & King-Shaw, E. (1998). *Canadian Tests of Basic Skills, Forms K and L.* Toronto: Nelson, A Thomson Company.

Hunter, C., & Robinson, S. (2002). *Literacy M.A.P. Meaningful applied phonics, explicit phonics through direct instruction.* Edmonton, AB: Edmonton Public Schools.

Hyde, J.S. (2005). The Gender Similarities Hypothesis. *American Psychologist, 60,* 581–592.

Jha, J., & Kelleher, F. (2006). *Boys' underachievement in education: An exploration in selected commonwealth countries.* London: Commonwealth Secretariat and Commonwealth of Learning.

Juel, C., Biancarosa, G., Coker, D., & Deffes, R. (2003). Walking with Rosie: A cautionary tale of early reading instruction. *Educational Leadership, 60,* 12–18.

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

Larsen, S. C., Hammill, D. D., & Moats, L. C. (1999). *Test of written spelling* (4th ed.). Austin, TX: Pro-Ed.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263–284.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21,* 159–173.

Mastropieri, M., & Scruggs, T. E. (1997) Best practices in promoting reading comprehension in students with learning disabilities. *Remedial and Special Education, 18*(4), 197–213.

Morrow, L. M., & Asbury, E. (2003). Current practices in early literacy development. In L. M. Morrow, L. B. Gambrell, & M. Pressley (Eds.), *Best practices in literacy instruction* (pp. 43–64). New York: Guilford Press.

National Institute of Child Health and Human Development U.S. (1999). *Teaching children to read: An evidence-based assessment of scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Reading Panel, US Department of Health and Wellness.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for reading instruction.* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office. [On-line]. Available at: http://www.nichd.nih.gov/publications/nrp/smallbook.cfm.

Pearson, P. D. (2001). Life in the radical middle: A personal apology for a balanced view of reading. In R. F. Flippo (Ed.), *Reading researchers in search of common ground* (pp. 78–83). Newark, DE: International Reading Association.

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283–319.

Phillips, L.M., Hayden, R., & Norris, S.P. (2006). *Family literacy matters: A longitudinal parent-child literacy intervention study.* Calgary, AB: Temeron/Detselig.

Phillips, L.M., Norris, S.P., & Mason, J. (1996). Longitudinal effects of early literacy concepts on reading achievement: A kindergarten intervention and five-year follow-up. *JLR: Journal of Literacy Research, 28*, 173–195.

Phillips, L.M., Norris, S.P., Osmond, W., & Maynard, A. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grade. *Journal of Educational Psychology, 94* (1), 3–13.

Phillips, L.M., Norris, S.P., & Steffler, D.J. (2005). *Meaningful applied phonics: A longitudinal early literacy study.* Vancouver, B.C.: Society for the Advancement of Excellence in Education.

PISA. (2005). *Measuring up: Canadian results of the OECD PISA Study.* Ottawa, ON: Statistics Canada.

Pressley, M. (1998). *Reading instruction that works.* New York: Guilford Press.

Pressley, M. (2002). *Reading instruction that works: The case for balanced teaching.* New York: Guilford Press.

Pressley, M., Allington, R. L., Wharton-McDonald, R., Block, C. C., & Morrow, L. M. (2001). *Learning to read: Lessons from exemplary first-grade classrooms.* New York: Guilford Press.

Snow, C. E., Burns, M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Stahl, S. A. (1992). Saying the "p" word: Nine guidelines for exemplary phonics instruction. *The Reading Teacher, 45,* 618–625.

Stahl, S. A., Duffy-Hester, A. M., & Stahl, K. A. D. (1998). Everything you wanted to know about phonics (but were afraid to ask). *Reading Research Quarterly, 33,* 338–355.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–406.

Strickland, D. S. (1998). *Teaching phonics today: A primer for educators.* Newark, DE: International Reading Association.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27*, 276–286.

Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33,* 321–363.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30,* 73–87.

Williams, J. (1995). Phonemic awareness. In T.L. Harris & R.E. Hodges (Eds.), *The literacy dictionary: The vocabulary of reading and writing* (pp. 185–186). Newark, DE: International Reading Association.

Williams, K. T. (2001). *Group reading and diagnostic evaluation (GRA+DE).* Circle Pines, MN: American Guidance Service.

How to cite this article

Phillips, L. M., Norris, S. P., & Steffler, D. J. (2007, April). Potential Risks to Reading Posed by High-Dose Phonics. *Journal of Applied Research on Learning, 1*, 1, Article 2, pp. 1-18