A Longitudinal
Early Literacy Study

Meaningful Applied Phonics

Linda M. Phillips
Stephen P. Norris
Dorothy J. Steffler

SOCIETY FOR THE ADVANCEMENT OF Excellence in Education

MEANINGFUL APPLIED PHONICS:

A LONGITUDINAL EARLY LITERACY STUDY

Meaningful Applied Phonics:

A Longitudinal Early Literacy Study

Linda M. Phillips, Stephen P. Norris, and Dorothy J. Steffler

May 2005, Research Series 23

SOCIETY FOR THE ADVANCEMENT OF EXCELLENCE IN EDUCATION

Library and Archives Canada Cataloguing in Publication

Phillips, Linda, 1952-
Meaningful applied phonics : a longitudinal early literacy study / Linda M. Phillips, Stephen P. Norris and Dorothy J. Steffler.

(SAEE research series ; #23)
Includes bibliographical references.
ISBN 0-9734046-9-8

1. English language—Phonetics—Study and teaching (Primary)—Alberta—Edmonton—Evaluation. 2. Reading—Alberta—Edmonton—Phonetic method—Evaluation. 3. Reading (Primary)—Ability testing—Alberta—Edmonton—Longitudinal studies. I. Norris, Stephen P. (Stephen Patrick), 1949- II. Steffler, Dorothy J. (Dorothy Jane), 1953- III. Society for the Advancement of Excellence in Education IV. Title. V. Series.

LB1525.3.P52 2005 372.45'5'09712334
C2005-902916-

SOCIETY FOR THE ADVANCEMENT OF EXCELLENCE IN EDUCATION (SAEE)

SAEE is an independent non-profit education research agency founded in 1996. The mission of the Society is to encourage excellence in public education through the provision of research to guide policy and practice. SAEE has commissioned 28 research studies to date and has a particular interest in examining innovations that may improve learning outcomes for less-advantaged students. As a registered Canadian charity, SAEE may provide official tax receipts for donations to its research.

For additional copies of this report, please contact:

SOCIETY FOR THE ADVANCEMENT OF EXCELLENCE IN EDUCATION (SAEE)
225 - 1889 Springfield Road, Kelowna, B.C. V1Y 5V5
Telephone (250) 717-1163 Fax (250) 717-1134 E-mail: info@saee.ca
Website http://www.saee.ca

ACKNOWLEDGMENTS

The authors wish to express their thanks to the children, principals, and teachers of the host schools for their help and collaboration, and the parents and families of the children for seeing the value of this study.

Grateful appreciation is also extended to Sandy Forster, Carolyn Hunter, Anne Mulgrew, Jean Ogonoski, Stuart Wachowicz, and Lorie Welk at Edmonton Public School Board for their help and collaboration.

We wish to acknowledge Jason Daniels, Alexandra Elton, Michele Fausak, Marianne Guilbert, Audrey Huculak, Lisa Hsu, Trudy Kwong, Shona Lazin, Alyson Letwin, Beth Mackey, Michelle Maier, Christine Marlowe, Agnes Maynard, Martha Smith, Sarah Spreen, Anna Tulinski, Debra Wayken, and Donna Wyatt for their assistance with data collection and scoring of tests. The contributions of Nabiha Rawdah in data collection, scoring, and data entry; Marion Leithead and Sandy Guilbert with the literature and program reviews; Randy Randhawa with the data analyses; Monda Blanche and Cherie Geering for personnel and payroll management; and Sharon Kanwal for checking on details are also gratefully recognized.

We also wish to recognize those whose financial support made this research possible: the Donner Canadian Foundation, an anonymous Alberta foundation, Edmonton Public Schools, Royal Bank Foundation, Teck-Cominco, HSBC, Coca Cola Canada, Investors Group, and the Society for the Advancement of Excellence in Education (SAEE).

We are indebted to the following who provided a peer review of the draft research report and offered thoughtful feedback to SAEE and to us: Dr. Marcia Barnes, Associate Professor and University Research Chair, Psychology Department, University of Guelph; Dr. Trevor Gambell, Professor of Curriculum Studies in the Faculty of Education, University of Saskatchewan; Dr. Graham Kelsey, Professor Emeritus, Faculty of Education, University of British Columbia, and Dr. Joseph Parsons, Department of Psychology, University of Victoria.

Finally, we extend our appreciation to SAEE Executive Director, Helen Raham, for her ongoing support over the duration of this undertaking.

ABOUT THE AUTHORS

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

Dr. 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.

Dr. Dorothy J. Steffler is Assistant Professor in Psychology at Concordia University College of Alberta and was the coordinator of the data collection for this study.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

EXECUTIVE SUMMARY

SCOPE OF THE STUDY

- The M*eaningful Applied Phonics* (M.A.P.) program is a prescriptive, explicit, teacher-directed approach designed to teach the alphabetic principle in a logical, sequential, and multi-sensory manner by teaching children the 70 graphemes that make up the 26 letters of the alphabet and strategies to segment and blend words into syllables.
- The seven schools using the M.A.P. program in this study were selected by the Edmonton Public School Board (EPSB).
- Students in the treatment group were taught using the M.A.P. program in Grades 1 to 3. These students were followed over the three grades and were tested at six points in time — in the fall and spring of each year until the end of Grade 3. The performance of these students was compared to that of control students also selected by EPSB at similar schools who did not do the M.A.P. program.
- This longitudinal study spans the years of 2001-2004.

MAIN FINDINGS

- Over the three years of the study, all seven treatment schools maintained their relative ranking with respect to the populations used to norm the tests.
- There was no systematic evidence to suggest that higher and lower fidelity of program implementation was related to differences in achievement.
- There was no systematic evidence to suggest that higher and lower M.A.P. method training was related to differences in achievement.
- SES was confounded with School such that its effects could not be distinguished from the effects of Schools.
- There was no systematic effect of Gender on achievement.
- The children in the Control group outperformed the children in the Treatment (M.A.P.) group.

CONCLUSION

- The evidence shows that children in the M.A.P. program performed less well than they would have performed had they been taught the programs used with the control children.

POLICY IMPLICATIONS

- Based upon the evidence produced in this study, continued use of the M.A.P. program is indefensible.
- Support for policy research such as this is critically important as a safeguard to ensure students receive effective programming.
- No program should be implemented without a thorough and systematic study of the published evidence of its effectiveness.

- All trials of innovative and experimental programs must be subjected to controlled comparisons and potentially disconfirmatory data as early as possible in the trials in order to protect children from the risks associated with less effective programs.
- Potential innovations must face strong critical appraisal in light of existing knowledge before they go to trial in the classroom.
- Adopting a policy of evidence-based practice is the surest guard against credulity.

1 Introduction to the Study

Effective reading instruction is a topic of abiding interest and debate. Many children who should be capable of reading well given adequate instruction are not doing so, suggesting that the instructional methods available to them are not appropriate (Pressley, 1998). In addition to the methods, there are a number of other factors that affect instruction such as SES, ESL, gender, and cultural differences. The prevalence of programs and methods dedicated to improvement of emergent reading instruction are numerous and varied. 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 study is *Literacy M.A.P. Meaningful applied phonics: Explicit phonics through direct instruction (*Hunter & Robinson, 2002). In 1998 and 1999 Edmonton Public School Board (EPSB) conducted two pilot studies of the precursor to the M.A.P. program. Although these studies provided data that seemed to suggest the precursor to the M.A.P. program "demonstrated significant increases in reading levels in grades one to four, beyond what would normally be expected" (Dosdall, 1999, p.2), the studies used a pretest and posttest only design and did not have control groups. Rather, comparison was made between the mean scores on the Highest Level of Achievement Test (HLAT) of the treated students and the district as a whole. This design and comparison did not provide a valid indication of how the treated children would have performed had they taken other programs because of its inability to eliminate most factors that jeopardize validity (Campbell & Stanley, 1963, pp. 5-12).

Thus, in collaboration with the Society for the Advancement of Excellence in Education (SAEE) and the EPSB, this three-year longitudinal study was proposed, planned and initiated over the 2000-2001 school year. Seven of the schools using the M.A.P. program were selected by EPSB to participate in this study. Data collection for the treatment group began with the Grade 1 students in September 2001 and continued each fall and spring until the end of Grade 3 in 2004. The research questions we were asked to address were the following:

1. Which schools achieved the strongest gains over three years?
2. Are higher and lower fidelity of method implementation and higher and lower quality of method training related to achievement gains?
3. Is SES related to achievement gains?
4. Is gender related to achievement gains?
5. Using nationally normed standardized scores, what was the gain made at each grade across all schools?

In addition to the comprehensive data collected on the students in the treatment program, performance scores on nationally and provincially normed standardized tests for the treatment group and for a random sample of control students matched on grade level, SES, gender, and school profile were procured in the Fall of 2004 and Winter of 2005. This report covers and analyzes all facets of this study in accord with the data collected for the treatment group, the data provided by EPSB for the treatment and control comparisons, and the EPSB-provided data on the fidelity of program implementation.

The following chapter provides a review of the relevant research literature on phonics, phonological awareness/phonemic awareness, and balanced reading programs. The M.A.P. program is described in chapter three. The methodology used in this study is the subject of the fourth chapter, followed by the analysis, results, and discussion in chapter five. Chapter six provides an analysis of the M.A.P program. The seventh and final chapter presents conclusions and recommendations.

2 Review of Relevant Literature

As a background to this study, it is helpful to examine what the research literature says about phonics and phonics instruction.

The ability to relate letters to speech sounds, knowledge that can be enhanced by instruction in phonics, plays an important role in reading (Adams, 1990; Pressley, 2002; Snow, Burns, & Griffin, 1998; Strickland, 1998). Phonetic ability, however, is but one component of successfully deriving meaning from print (Adams, 1990; Pressley, 2002; Snow et al., 1998; Strickland, 1998). The phonics debate centres not on whether phonics is necessary or not, but on its importance for all students relative to other strategies to construct meaning from print (Flippo, 2001; Goodman, 1993; Pressley, 2002; Rasinski & Padak, 2004).

Research to date does not provide definitive answers about how much emphasis should be placed on phonics, nor about the best way to teach phonics (Cunningham, 2003; National Institute of Child Health and Human Development [NICHHD], 2000; Stahl, Duffy-Hester, & Stahl, 1998). Thus, the debate continues. The following sections review the research on phonological and phonemic awareness, phonics instruction, and balanced reading programs. Based on this review, an emerging consensus can be discerned, one potentially useful in planning effective reading instruction for beginning readers.

PHONOLOGICAL AWARENESS/PHONEMIC AWARENESS

Phonological awareness and phonemic awareness both more or less refer to awareness of the sounds (phonemes) that make up spoken words. Phonological awareness, the broader term, is awareness of the constituent sounds of words in learning to read and

spell. The constituents of words can be distinguished in three ways: a. by syllables, as /b□k/, b. by onsets and rimes, as /b/ and /□k/ , and c. by phonemes, /b/ and /□/ and /k/ (Harris & Hodges, 1995, p. 187). Phonemic awareness focuses on the ability to differentiate and manipulate individual phonemes represented by letters and letter clusters within words (Williams, 1995, p. 185). Following the example of Blachman (2000), Gillon (2004), and Snow et al. (1998), we use the more inclusive term phonological awareness, considering phonemic awareness to be a subset of this skill. Measures used to assess phonological awareness (PA) include tasks evaluating children's ability to identify phonemes in words, to divide words into onset and rime segments, to categorize phonemes, to segment a word into phonemes, to blend phonemes into words, and to manipulate phonemes in words.

Numerous studies of PA conducted individually and collectively over the past three decades indicate a strong positive correlation between young children's phonological awareness and their ability to learn to read. Studies demonstrating a strong and positive correlation between PA and reading ability are many (Bradley & Bryant 1983; Byrne & Fielding-Barnsley, 1995; Lundberg, Frost, & Petersen, 1988; Lundberg, Olofsson, & Wall, 1980; Perfetti, Beck, Bell, & Hughes, 1987; Torgeson, Wagner, & Rashott, 1994; Vellutino & Scanlon, 1987). Data collected in such studies has led some reading researchers to conclude that PA has a causal relationship to reading (Blachman, 2000; Bradley & Bryant, 1983; Bus & van IJzendoorn, 1999; Vellutino & Scanlon, 1987; Wagner, Torgesen, & Rashotte, 1994). Others maintain that although PA plays an important role in learning to read, causality has not yet been established. Gillon (2004), for example, states that PA is necessary for reading and spelling success, but cautions that PA is not by itself sufficient for successful reading performance.

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; Wagner et al., 1994). In explaining this reciprocity, Perfetti et al. (1987) hypothesized that learning to read increases reflective phonemic awareness which then promotes further gains in reading ability. Research indicating that children who receive a combination of PA and letter-sound training score higher in posttests assessing PA, reading, and spelling (e.g., Schneider, Roth, & Ennemoser, 2000) contributes to the recommendation that letter training be included in PA programs (Bus & van IJzendoorn, 1999; Ehri et al., 2001). Other researchers such as Morris, Bloodgood, Lomax, and Perney (2003) conclude that concept of word in print follows alphabet knowledge and beginning consonant awareness, leading them to call for early teaching of the alphabet and beginning consonant sounds.

A sequence of phonological sensitivity, defined by Anthony, Lonigan, Driscoll, Phillips, and Burgess (2003) as "sensitivity to and the ability to manipulate the sound structure of oral language" (p. 473), has been proposed by Adams (1990), Anthony et al., (2003),

and Goswami and Bryant (1990), among others. Anthony et al. identify the sequence of sensitivity to move from mastery of word level differentiation skills to syllable-level skills, to onset/rime skills, followed by phoneme-level skills, the same sequence outlined by Goswami in 2000. Anthony et al. additionally conclude that within this developmental sequence, stages are overlapping rather than temporally discrete. Like Adams (1990), Anthony et al. report that blending skills emerge before segmenting skills when linguistic complexity and cognitive operations are controlled.

Based on these findings, Gillon (2004), and Snow et al. (1998) have outlined an instructional PA sequence. Snow et al. 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 "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).

Implications for instruction emerge out of the research studies on PA and the meta-analyses of these studies. First, 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 (2000), wrote that both kindergarteners and preschoolers gain most from such training, surmising that this is because they start out with the least phonological awareness. The NRP, which also included a length-of-instruction variable in its meta-analysis, concluded that 5 to 18 hours of PA instruction was more effective than that lasting a shorter or a longer time. Ehri et al. (2001) wrote, "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 "Training 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 variances in reading skills. Ehri et al. (2001) calculated 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. They note, in addition, that PA should be viewed as a means rather than an end and, like Bus and van IJzendoorn, caution that PA is not the sole key to reading success.

Analyses of research have failed to identify PA instructional programs that appear significantly more effective in helping students learn to read than others offering systematic PA instruction, although commercial programs are seldom examined in the research literature (Shanahan, 2003). Instead, the NRP notes that a variety of programs was found to be effective, although programs that concentrated on developing one or two phonemic awareness skills (such as segmentation and blending) were found to be

more effective than programs teaching three or more skills. The NRP analysis also concluded that PA instruction carried out in small groups was more effective than individual or whole group instruction.

It has been suggested that the order in which children generally acquire PA offers a guide to planning an instructional sequence (Adams, 1990; Anthony et al., 2003). Using this concept of order, Gillon (2004) suggests that instruction proceed from the development of students' awareness of speech sounds to word-syllable-letter awareness, onset-rime awareness, and then to phoneme skills such as blending, segmenting, and manipulating. In addition, Gillon provides a guide to adjusting task difficulty for each of the following PA skills: syllable segmentation, rhyme, phoneme identity, blending, segmentation, sound deletion tasks, identifying and describing phoneme changes, and matching phonological form to orthographic form.

Individual student differences should be taken into account when planning PA instruction (e.g., Blachman, 2000; Ehri et al., 2001; Gillon, 2004). As some children come to reading already aware that speech can be segmented and have a rudimentary ability to manipulate phonemes, while others lack this awareness and skill, differentiated instruction is required if all students are to receive appropriately paced reading instruction.

It is important to note that the NRP members only claim that "Phonological analysis instruction is more effective than alternative forms of instruction or no instruction in teaching phonological analysis" (Ehri et al., 2001, p. 274). That is, after receiving PA skill training, students were found to be better able to read words using the PA skill in which they had been instructed. Voicing his skepticism of many of the claims made for PA, Scholes (1998) writes that when reading is defined as understanding written text, rather than sounding out words, PA "has very little, if anything, to do with reading or the acquisition of literacy" (p. 178). In the NRP meta-analysis, although PA instruction was shown to teach PA very effectively, such instruction had only a small-to-moderate impact on reading comprehension (Ehri et al., 2001). Thus, while PA has been shown to be important, perhaps crucial, in reading words, it should be remembered that PA is but one component among a complex array of literacy experiences that help children learn to read (Cunningham 2001; Pressley et al., 2001, Snow et al., 1998).

WHAT IS 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, that letters can represent sounds, as well as knowledge of the sounds represented by those letters or letter combinations. Phonics is a way of teaching reading and spelling that stresses symbol-sound relationships, used especially in beginning

instruction (Harris & Hodges, 1995, p. 186).

Although there is no disagreement with the proposition that all readers of English use phonetic knowledge in making sense of print (Adams, 1990; Flippo, 2001; Goodman, 1993), the centrality of that knowledge for text comprehension is contested. As well, the kind of phonics instruction indicated by the descriptors systematic, explicit, and intensive is a source of controversy (and there are at least six other types). Furthermore, there is often heated disagreement regarding approaches to phonics instruction, particularly between those advocating a synthetic approach and those favouring an embedded approach. The former approach, associated with traditional phonics instruction, adheres to a parts-to-whole sequence where teaching proceeds from instruction in identifying the symbols and sounds for letters and groups of letters to blending these letters together to form words (Chew, 1997; Morrow, Holt, & Sass, 2002; Stahl, Duffy-Hester, & Stahl, 1998; Strickland, 1998). In the latter, referred to as a more contemporary approach (Stahl et al., 1998), phonics instruction occurs in the context of reading and writing, moving from consideration of whole text to an examination of the phrases, words, and word parts necessary for text comprehension (Strickland, 1998).

A lack of research evidence indicating the superiority of any one approach to teaching phonics (Mazzoni & Gambrell, 2003; NICHHD, 2000; Stahl et al., 1998) hampers resolution of this debate. Similar to Shanahan's observation (2003) regarding a paucity of research comparing commercial PA programs, Cunningham and Cunningham (2002) note that few instructional studies have compared different types of phonics instruction and those that have, have often compared systematic phonics instruction with little or no instruction, or with random phonics instruction. In such comparisons, systematic phonics instruction has been shown to be beneficial for promoting reading ability (see Bus & van IJzendoorn, 1999; Ehri et al., 2001).

Despite differences of opinion, there are a number of propositions concerning phonics instruction that have received widespread support. First, there is agreement that learning to read is not easy; indeed, it is a highly complex process (Adams, 1990; Flippo, 2001; Goodman, 1993; Snow et al., 1998). Spache and Berg (1955) wrote, "Reading is more than going rapidly over the lines of print, and it is more than slowly and laboriously plodding along looking carefully at every symbol and form. Rather, reading is a complex of several skills which must function together to produce an amount of comprehension in keeping with the purposes, needs, and methods of the reader" (pp. 3-4). Second, as learning to read is not a simple task, reading instruction must develop the variety of skills necessary for successfully constructing meaning from text (Cunningham, 2003; Pressley, 2002; Snow et al., 1998; Strickland, 1998). Moreover, the NRP states, it must be remembered that phonics instruction is a means to an end, that end being reading comprehension. Third, similar to teaching PA, no single method of phonics instruction has been shown to be the best way to teach this skill.

There are, however, a number of practices that receive recognition in the field of reading research. It appears, for example, that an early introduction to letter-sound correspondences helps children to learn to read words. The NRP meta-analysis determined that phonics instruction is most effective when introduced in kindergarten and grade one. In a study of four classrooms, Juel and Minden-Cupp (2000) found "phonics first and fast" helped low-reading group children make substantial gains in successful word reading over the course of grade one. "First" because these children spent more than one-third of their language arts time in the fall engaged in phonics activities. "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. However, Stahl also wrote that although there is little research indicating an optimal amount of time for phonics instruction, he would recommend that no more than 25% of reading time should be spent on phonics instruction and practice, a percentage also advocated by Cunningham and Cunningham (2002). And despite Juel and Minden-Cupp finding that intensive phonics instruction helped low-reading group students make substantial gains, Juel, Biancarosa, Coker and Deffres (2003) advise that preoccupation with teaching decoding skills may have a negative impact on students' overall reading ability — that decoding, although crucial, is not sufficient.

Good phonics instruction focuses students' attention on the letter-sound correspondences in words, thus helping them to develop an awareness of the internal structure of words and the skill to detect patterns useful for recognizing words efficiently (Adams, 1990; Snow et al., 1998; Stahl, 1992; Stahl et al., 1998). Such instruction is often referred to as explicit. As Adams (1990) wrote, when children sound out words, they engage in careful study of those words. And through a continual study of words, spelling patterns are revealed which help children build orthographic automaticity. She, however, cautioned that too many pauses to study words when reading text will hamper text comprehension, the goal of reading. Stahl (1992) also advocated phonics instruction that forces children to look closely at the internal structure of words, at the patterns in words, because "it is through the learning of these patterns that children learn to recognize words efficiently" (p. 624). As mentioned, it is not clear from the research reported to date whether synthetic phonics or embedded phonics is the better method to teach children to read. Stahl et al. (1998) recommended both, writing that good phonics instruction provides practice reading words both in isolation, where words are more likely to be examined closely, and in stories, where phonics knowledge can be strengthened through application while helping students recognize the value of phonics for comprehending text. Stahl et al. (1998) hypothesized that if a method is able to help children construct knowledge about words, it may not matter much just how it is done. In addition, they noted that good phonics instruction focuses on words, not rules.

Systematic instruction is widely advocated (Morrow & Asbury, 2003; NICHHD, 2000; Pressley, 2002; Strickland, 1998). Avoiding the term systematic, which has become highly politicized, Morrow and Asbury (2003) refer to the need for organized and comprehensive skills development, while Strickland (1998) writes that "instruction is systematic when it is planned, deliberate in application, and proceeds in an orderly manner" (p. 51). Cunningham and Cunningham (2002) also point out that although children need to learn how to decode sequentially, this does not indicate that instruction should be through synthetic phonics lessons. The call for systematic phonics suffers, too, from a lack of evidence of a particularly effective system. For example, phonics instruction has often moved from instruction in consonant letter-sound correspondences, to the vowels – focusing first on short, then on long vowel sounds. However, based on a reanalysis of a phoneme-grapheme frequency study, Fry (2004) suggests that it makes more sense to introduce consonants and vowels by their frequency of use. 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 can support.

In addition to questions about a systematic ordering of phonics instruction, there is growing evidence that some children have problems using analogy as a decoding strategy, while it is the preferred strategy for other children (Pressley, 2002). Juel and Minden-Cupp (2000) report that some poor readers in their study had problems seeing related letter clusters (variously referred to as rimes, word families, or chunks). Given the limited research data available on these print processing differences, it is probably safe to accept Juel and Minden-Cupp's comment that the phonics program in their study that included both onset-rime and sounding out and blending phonemes within the rimes seemed very effective – that it is not an either/or proposition.

Writing using invented spelling is another widely recommended practice (Adams, 1990; Craig, 2003; Richgels, 2001; Snow et al., 1998; Stahl, 1992). Adams holds that the process of inventing spelling is essentially a process of phonics. That is, when children must break down the words they want to write into individual phonemes and then have to attach a letter to each of these phonemes, they are actively involved in phonemic segmentation tasks. Gillon (2004) suggests also that spelling activities promoting the use of explicit phonological processing knowledge play an important role in early reading development.

Additionally, the teacher in the Juel and Minden-Cupp (2000) study who engaged her low reading group in intensive phonics instruction also offered quite different instruction to her other two reading groups. Because it is apparent that not all students benefit equally from one set of instruction, that one size does not fit all, differentiating instruction is another widely supported practice (Goldenberg, 2001; Pressley, 2002; Pressley, Allington, Wharton-McDonald, Block, & Morrow, 2001; Snow et al., 1998). In their report, the NRP recommends flexible pacing and flexible grouping of students by phonetic

ability, and the International Reading Association (IRA) (1997) states that "effective teachers of reading and writing ask when, how, how much, and under what circumstances phonics should be taught" (p.4). This statement brings us back to a final widely-accepted proposition concerning phonics instruction – the important role teachers play in the development of children's reading ability.

Skillful, knowledgeable, dedicated teachers who are able to identify and implement instructional approaches and practices as the situation demands are recognized as the key to reading success (Duffy & Hoffman, 1999; Flippo, 2001; Juel & Minden-Cupp, 2000; Mazzoni & Gambrell, 2003). Similarly, an IRA (1999) position paper contends that educators familiar with a wide range of methodologies are in the best position to make decisions regarding the most appropriate instruction for their students. And Snow et al. (1998) write, "If we have learned anything from this effort [reviewing research on reading development and instruction], it is that effective teachers are able to craft a special mix of instructional ingredients for every child they work with" (p. 2-3). Research provides data regarding the importance of such skills as phonological awareness and phonics for reading and how children may develop these skills, but it is teachers who use this knowledge in their interactions with students. It is the teachers who craft programs to help specific children learn to read. And, increasingly, reading experts are emphasizing the need for these programs to balance skills teaching and holistic literacy experiences (Pressley, 2001).

WHY ARE BALANCED PROGRAMS IN READING NECESSARY?

After presenting, in detail, 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 et al., 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 competencies 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). Others share P. David Pearson's view and they include Morrow, Gambrell and Pressley (2003), Strickland (1998), and Torgesen (1998). Pressley (2002) succinctly describes his envisioned balance as one "including prominent skills instructions and prominent holistic reading and writing experiences" (p. 170).

In addition to planning a balanced reading program for their classrooms, teachers also need to provide balanced reading instruction to individuals so all students may develop the ability to comprehend written material at a level consistent with their general ability and knowledge (Hoffman, 2000; Torgesen, 1998). Matching instruction to children's needs will often entail moving those who understand and are able to use the alphabetic principle on to more advanced reading activities (Snow et al, 1998). Reading instruction balanced for student success will also reflect the understanding that students at risk for reading difficulties generally require more explicit and comprehensive reading instruction than normally progressing students (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Juel & Minden-Cupp, 2000; Schneider et al, 2000; Torgesen, 1998). In concluding their discussion, Foorman et al. (1998) remind readers that the direct instruction in letter-sound correspondences they found to be particularly effective for children at-risk for reading failure took place in print-rich environments with a significant literature base; that is, there was balance in the total reading program for these children.

SUMMARY

From the research and interpretations of research cited, a broad consensus regarding phonological awareness and phonics instruction emerges. First, it is widely accepted that the ability to distinguish the phonological components of speech and to manipulate sound and letter correspondences is related to reading success. These skills appear to be developed most efficiently through explicit instruction that is organized and comprehensive. However, no one "best way" has been shown to help all children develop these skills. Furthermore, children come to school with very different background experiences and abilities. Given these factors, teachers need to differentiate instruction if all their students are to have a fair chance for appropriate growth in their ability to read. And because phonological awareness and phonics skill development is just one component of learning to read, successful reading programs incorporate skill development into an instructional program that involves abundant practice reading and being read to from quality literature. Finally, as the NRP writes, it must be remembered that phonics instruction is a means to an end, that being the ability not just to decode but to comprehend written text.

3 Description of the Meaningful Applied Phonics Program

Literacy M.A.P.: Meaningful Applied Phonics, Explicit Phonics through Direct Instruction (hereafter referenced as M.A.P.) is published and distributed by Edmonton Public School Board (Hunter & Robinson, 2002). The M.A.P. authors advocate teacher-directed whole class instruction and a sequential development of skills. 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.

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 refereed 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, ô, oo, and □). 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 a 1, 2, 3, or 4. That is, an ordinal number is placed above a grapheme using its 2nd, 3rd, 4th, 5th, or 6th sound. Students are asked what sound they hear first, second, etc. (p. 18) and must bear in mind that every syllable must have a vowel, that vowels can have more than one sound, and only vowels can say their names in a word (p. 51), and then there are five spellings of 'er', five jobs for silent 'e', and there are 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 2 over the "th" to indicate that the second sound is like the "th" in "this" (see page 105). Other examples are provided in Appendix E.

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 sentences not begin until students have been taught the first 54 graphemes, the five jobs for silent 'e', and three weeks of dictated sentences. By way of example, here is a paragraph about **Me:** 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. Graphemes taught in isolation begin the reading process (see Appendix A). Graphemes are followed by spelling words, dictated sentences, dictated paragraphs and stories, reproducible stories, and then children's literature. The M.A.P. program also incorporates instructions and strategies for teaching fluency, comprehension, and vocabulary. For example, in the section titled "Comprehension Activities," pre-reading discussions, prediction, 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, 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" (p. 4). Thus, the program manual, like the teaching program, provides explicit directions intended to help teachers to direct children to reading success.

More specifically, the M.A.P. resource is 311 pages in length which includes program explanations and some 200 pages of graphemes, spelling word lists, grapheme sequence lists (115 pages), and thematic word lists. A glance at the Table of Contents provides an overview of M.A.P. and includes topics such as "Teacher Directed Whole Group Instruction" and "Sequence of Instruction (p. 5), "Teaching the Graphemes" (p. 14), "Spelling Process" (p. 19), and "The Writing Process" (p. 21), "The Reading Process" (pp. 27-28), "Vocabulary Instruction Strategies" (p. 32), "Comprehension Activities" (pp. 33-41) with graphemes and new spelling words (30 mins) combined with reading (60 mins) into a sample timetable for 90 minutes each day (p. 42), itemized lesson plans (pp. 43-48), colour-coded paragraph models (p. 49), extensive organization sequence charts for consonants, vowels, "Five Jobs for Final Silent e" (pp. 55, 69), "Five Spellings of er" (pp. 56, 70), nouns (pp. 59, 73), verbs (p. 60), synonyms and antonyms (p. 61), spelling rules (pp. 81-84), syllabication rules (p. 85), school to home communications in the form of "Dear Families" letters (pp. 86-90), charts for graphemes (pp. 91-101), and spelling lists (pp. 102-103) accompanied by a "Sample Sequence Grapheme/Spelling Program for each grade level (pp. 105-220), Thematic Word Lists (pp. 221-230), and a "Creative Writing Word List" (pp. 230-237). The Thematic Word List divides the words into topics such as the zoo, days of the week, months of the year, school, plants, mystery, and so on, for students to use while creating stories. The Division II (Grades 4-6) contains approximately 59 pages of word lists (pp. 238-291), presented in much the same manner as those pages for Division I students (K-3).

Hunter and Robinson (2002) maintain that learning how to "decode words quickly" will "make reading easier and more enjoyable" (p. 7). They allude to decades of "debate" regarding "a phonics based approach as opposed to whole language methods," adding that "It is time to make children the focus." (p. 3). The M.A.P. phonics approach is described as "multi-sensory" in that students are asked to see the printed grapheme, say it, hear it, and write it (p. 5), as well as tap "out the syllables" (p. 8). While M.A.P. is based on and reflects a teaching philosophy and methods that were developed almost fifty years ago (p. 4), the authors refer also to current research reported by the National Institute of Child Health and Human Development (NICHHD) (1999) to report that children learn to read best when:

- they are taught and understand the alphabetical principle of the English language;
- systematic sequential [whole group] instruction in phoneme awareness and phonics is explicitly taught;
- they are taught how to apply their phoneme awareness and phonics to spelling and reading;
- programs are a means to an end;
- programs include instruction on phoneme awareness, phonics, fluency, vocabulary, and comprehension" (p. 3).

Familiar terms such as "scaffolding" and "active participation" are defined by M.A.P. respectively as reducing complex tasks to smaller more manageable steps, and having students "see it, say it, hear it, write it" (p. 5). M.A.P.'s "check for understanding", one of seven instructional strategies, consists of frequent spelling tests, is included in a Table, (see Appendix B), and is followed by "guided practice" (using grapheme cards), systematic "immediate feedback" during seatwork, and "time and opportunity to apply new knowledge" in "independent practice".

The "Explanation for Students" (p. 7) provides a rationalization for phonic awareness and instruction, spelling instruction, reading fluency, reading comprehension, grammar instruction, writing instruction – not however, in the sequence presented in the M.A.P. manual. The reading process in the M.A.P. phonetics sequence of instruction is, as indicated earlier, presented last (pp. 27-28). M.A.P. seems to equate decoding with reading and claims that decoding words "quickly and smoothly" will make reading easier and more enjoyable (p. 7). By the time the students have advanced to Division II (Grades 4-6), they should be "familiar with 19 spelling rules, the markings, and the rules of syllabication," and be finding "there are irregularities in the English language" (p. 238).

Children's literature is the last step in the chain that links graphemes to spelling words, to sight words, or dictated sentences, dictated paragraphs and stories, reproducible stories, and then to children's literature in the "Reading Process" model (see Appendix C). M.A.P. twice suggests the incorporation of reading or other reading instructions, into phonics instructions. First, citing a summary from the NRP Report, M.A.P. states, "It is important to emphasize that systematic phonics instruction should be integrated with other reading instruction to create a balanced reading program. Phonics should not become the dominant component in a reading program, neither in the amount of time devoted to it nor in the significance attached to it" (p. 11). Later, in a section regarding "Indirect Instruction" (p. 32), M.A.P. again cites the NICHHD Report (1999, 4-17) as stating, "students are exposed to words or given opportunities to do a great deal of reading, or listen to read alouds." However, when quantifying actual time spent reading, the M.A.P. time element must be kept in perspective, as 90 minutes are allotted each day to phonics-related exercises in the sample M.A.P. timetable (see Appendix D).

The M.A.P. program also prescribes writing-related activities such as blocking strategies and dictated paragraphs (p. 22). Blocking, for example, entails stopping to demonstrate each word in a sentence such as "We can play ball," by "holding both hands out to form a brick or block shape," moving the hands "from left to right as if carrying a block," and placing the blocks side by side in the sentence. Dictated paragraphs, also to be "presented to the students in small sequential steps," can only be undertaken once students have been taught the "first 54 graphemes," the "five jobs for final silent e," and have experienced "three weeks of dictated sentences" (p. 22).

Several M.A.P. instruction strategies regarding students' writing-related activities,

"Dictating Stories," for example, include teacher-chosen picture prompts "will provide ideas for their [the students'] stories." Moreover, the thought process for these stories is "modeled" or prescribed in a six-step process: "look at the picture, think the first sentence, say it, print and punctuate it, back up and check, then think the next sentence, and …" (p. 25).

The Division II (Grades 4-6) portion of the M.A.P. program manual (pp. 238-310), uses a similar approach to that of Division I. Division II provides 66 pages of charts and word lists (easier words learned in previous grades, combined words or multiword terms, prefixes, roots, suffixes, homonyms, homographs, homophones, more/different/new words), prefaced by a two-page note that indicates lists are designed principally for students in Grades 4-6 studying the Alberta curriculum in Social Studies, Science, and Mathematics.

M.A.P. points out those Grade 4 students will find "that there are "irregularities in the English language," in that "spelling rules do not apply, the markings do not necessarily fit, and the dictionaries do not always agree on where syllable divisions should be" (p. 238). The spelling anomalies are explained by a brief history of the "Evolution of the English Language" and an example of "Linguistic Trivia" (pp. 238-240). A one-page resource list provides "Resources to Enhance Teacher Background" (p. 243) which consists of ten resources regarding prefixes, roots, suffixes, teaching spelling, the act of spelling, and three dictionaries. The Division II section concludes, "Above all, have fun! Make the lists your own" (p. 242).

It is this program, introduced in selected EPSB schools for Grades 1-3, whose effectiveness is the focus of this research. The following chapter provides an overview of the research method.

4 Method

There are several difficulties in conducting research on any program or method where neither the treatment students nor their schools were picked at random. Students in the control group were matched to the treatment group on the basis of school size and population similarity.

DESIGN AND SAMPLE

The study compared seven treatment groups by school to a control/comparison group identified by EPSB. The control group was given no treatment other than the program in use in the schools at the time of the study. Pretests and posttests were given to all treatment groups and EPSB and provincial results were used as pretests for the control groups in Grade 1 and as posttests in Grades 2 and 3 (see untreated, non-equivalent control group design with pretests and posttests in Cook and Campbell, 1979, pp. 103-118).

Grade 1 students from seven schools in the EPSB participated in pre- and post-test data collection at six points in time across the three-year longitudinal study: Grade 1 (Fall 2001 and Spring 2002); Grade 2 (Fall 2002 and Spring 2003); and Grade 3(Fall 2003 and Spring 2004). Multiple dependent measures, administered by EPSB, were used as posttests to reduce the risk of systematic bias.

Seven schools using the M.A.P. program were identified by EPSB in the winter of 2001. Six of the schools are urban and one rural. The six urban schools represented a range of multi-cultural 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% in the case of Schools 6 and 7. School 1 served an Indigenous population of children. The multi-cultural diversity of School 6 is predominantly East European. In the remainder of schools the population is predominantly Asian. Mobility rates for the years 1999 to 2004 vary 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 mobility data were available for School 1.

The control group was selected by EPSB to match only the students for whom complete data was available from Grades 1-3 inclusive. Control group students were selected to match as closely as possible the treatment students on gender, school size, SES, ESL, and cultural diversity. The control group data were not made available until the spring following final data collection on the treatment group.

At the beginning of Grade 1, in the fall of 2001, 272 students were registered across the seven schools. The range in attendance during the fall testing period on any given testing day was from 209 to 252 students. Complete data were collected for only 209 students which represents an absentee rate of approximately 33% — one-third of the Grade 1 sample during the three-week testing period. The Grade 1 students were followed until the end of Grade 3. By the end of Grade 2 in the spring of 2003, the sample for which complete data were available had dropped to 170, an attrition rate of 37%. By the end of Grade 3 in the spring of 2004 the treatment group population for which complete data were available had dropped to 137, an attrition rate of approximately 50% across the seven schools over the six testing periods across the three years. These decreases in numbers does not necessarily mean that students actually left the schools. Rather, it means that they were not available for some aspect of the testing.

Only students for whom complete data were available from Grades 1, 2, and 3 were included in the treatment versus control group comparison. Within each grade, only students who were part of the study from the outset and for whom we had complete data for the specific testing for the specific grade were included in the analyses. For instance, if a student missed only one subtest, then we did not have complete data for that student, and hence the student was not included in the analyses. Bear in mind that the attrition rate reflects two factors: the reason just cited and transfer to another school.

ASSESSMENT INSTRUMENTS

The following measures of students' literacy development were used in the study:

Group Reading Assessment and Diagnostic Evaluation *(GRA⁺DE)*. Reading performance was measured in the fall and in the spring 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. Form A was used for fall testing and Form B for spring testing. Grade 1 children 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, Same and Different Words, Print Awareness, Letter Recognition, Phoneme-Grapheme Correspondence, Listening

Comprehension, and Word Reading. All subtests are multiple-choice format with four options.

The ***Sound Matching*** subtest is a measure of phonological awareness of beginning 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).

The ***Same and Different*** *Wor*ds 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 Awarene*ss** 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).

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.

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).

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 horse is jumping over the fence" (Williams, 2001).

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).

The two subtests used from the Level 1 GR*A*$^{+}$DE included Word Meaning and Sentence 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.

Children in Grade 2 were tested using the Level 2 GRA+DE including the following subtests: Listening Comprehension, Word Reading, Sentence Comprehension, Word Meaning, and Passage Comprehension. All subtests are multiple-choice format with four options. The *Listening Comprehension, Word Reading, Sentence Comprehension*, and *Word Meaning* subtests are similar in format and administration to the previous explanations. *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 the content of the passage.

Grade 3 children were tested using the Level 3 GRA+DE including the following subtests: Listening Comprehension, Word Reading, Sentence Comprehension, Vocabulary, and Passage Comprehension. All subtests are multiple-choice format with four options. The *Listening Comprehension, Word Reading, Sentence Comprehension*, and *Passage Comprehension* subtests are similar in format and administration to the previous explanations. The *Vocabulary* subtest is a measure 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, for example, 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, 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.

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 EPSB, who receive instruction in English, wrote the Highest Level of Achievement Tests (HLAT) in reading and writing. For the purposes of this study, only the reading subtest of the HLATS which consisted of the reading comprehension subtest of the *Canadian Test of Basic Skills* (Hieronymus, Hoover, Lindquist, Scannell, King-Shaw, 1998) was used. The HLAT testing program provides student, school and district information that is different from that provided by the provincial achievement testing program.

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 of the PAT is of interest here). The purpose of the achievement tests is to determine whether students are learning what they are expected to learn, to inform Albertans about students' achievement relative to provincial standards, and to assist schools, school authorities, and the province in monitoring and improving student learning. At the end of May, students in Grade 3 take a 70-minute writing test and at the end of June a 60-minute reading test. These tests are a combination of multiple-choice and numerical-response questions, and written response sections are scored by select classroom teachers and are centrally scored at Alberta Education.

TEACHER SURVEY ON FIDELITY OF IMPLEMENTATION AND TEACHER PREPARATION IN M.A.P

A teacher survey was designed, developed, and administered by EPSB personnel 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 an EPSB personnel. Survey responses from 37 teachers were collected by EPSB personnel and sent to us. An independent rater judged teachers' qualitative and quantitative responses. The sections focused on the phonograms and spelling were each rated on a scale from 0 to 10. The four remaining sections dealing with number of spelling words per week, reproducible charts, homework, and method training were each rated on a scale from 0 to 3. The highest score on each scale indicated the highest level of implementation. The total implementation score for each school was obtained. These scores provide a comparative ranking only of implementation levels. From these scores it is not warranted to infer that a school with the highest score on a category (say 3) was implementing the program perfectly with no possibility of improvement, or that a school with the lowest score on a category (say 0) was not implementing the program at all. (For a copy of the teacher

survey, see Appendix F.) The range of implementation across teachers and schools falls short of full implementation by all teachers and schools that would have provided the most robust test of the M.A.P. program. However, the range observed has the strength that it likely better represents the levels of implementation that would be found in any widespread use of the M.A.P. Teachers of all subject areas are known to adapt programs to varying degrees to match the different circumstances they encounter.

Program Elements

The fidelity of implementation survey included seven main categories of items. These included the following questions: (a) the number of phonograms used/taught (a graphic character or symbol represent a phonetic sound, phoneme, or word), (b) the amount of time allotted each day or week for phonograms, (c) number of spelling words used/ taught, (d) the amount of time allotted each day or week for spelling, (e) reproducible charts used/taught, and (f) whether there was a homework routine. Teachers were invited to provide additional comments.

Program Training

This 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.

PROCEDURE

The instruction methods used in the treatment and control groups are briefly summarized below:

Treatment

The treatment group was taught using the *Literacy M.A.P. Meaningful applied phonics: Explicit phonics through direct instruction (*Hunter & Robinson, 2002) program as outlined and discussed previously. The same group of students was followed for three years starting at the outset of Grade 1 in the Fall of 2001 and completing in the Spring of 2004. For a more thorough description of the M.A.P., see Chapter three. The essential features of M.A.P. as presented by Hunter and Robinson include:

- M.A.P. is an explicit, teacher-directed approach;
- 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,
- M.A.P. segments and blends words into syllables when reading and writing unfamiliar words,
- M.A.P. teaches the 70 graphemes for the first 26 single letters sounds of the alphabet as well as vowel and consonant digraphs,

- M.A.P. focuses on phonics, spelling and grammar for approximately 30 minutes each day combined with 60 minutes of reading and writing.

Control

The comparison children were taught a balanced literacy approach modeled after Patricia Cunningham and Dorothy Hall's *Four-Blocks Literacy Model* (2001). 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) and was developed with the needs of diverse students (literacy levels, interests, and ways of learning) in mind.

Students are not taught using Four-Blocks instruction until Grade 1. Starting in Grade 1 (the program has two components – Grades 1-3 and Grades 4-6+ - only the former 1-3 is relevant to this report), literacy instruction is divided into four sections, each requiring approximately the same length of 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 text, "is always focused on comprehension" (see website). Big books and the initial reading texts are read together by the teacher and class. Teachers introduce other books and formats as the year progresses, work with the whole class on occasion, but students also start to read together in small groups, with partners, individually, or with the teacher. During the Self-Selected Reading Block, children read what they themselves have chosen. Students are asked, in addition, to respond to what they have been reading, sometimes sharing their reading and responses with others, and conferencing with the teacher.

The Writing Block includes mini-lessons where students learn the fundamentals of writing, for example, 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 as they proceed through the writing process. 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" (see website). Strategies include Making Words and Word Wall with suggestions for grouping such activities into a 5-lesson cycle. This approach is intended to help teachers meet mandates for including systematic, sequential phonics in their reading instruction. Information taken from: www.four-blocks.com (2001, Carson-Dellosa Publishing Company available as cited on 11 May 2005).

A personal communication with Jean Ogonoski at Edmonton Public School Board (6 May 2005), indicated that the four blocks are collapsed into 3 large blocks: Working with Words, Reading, and Writing. These were outlined by her as follows:

1. Working with Words (35 minutes)
    - Word Wall (10 min) - works on high frequency words; introduces 5

new words per week with challenge words for advanced students; reviews words every 3-4 weeks; practice task sent home every Monday night (125 core words per grade 1, 2, 3)

- Making Words and other Word Patterns (25 min) - whole class lesson with students building words with letter cards on desks; sorting and transfer word tasks; and possibly a home activity every few weeks (words here are independent of core words used in word wall)

2. Reading (70 minutes)
    - Read Aloud (15 min) - model expert reading; expose students to sounds and sense of written language forms; encourage enjoyment of written language
    - Shared Reading (15 min) - enables all students to participate in reading (e.g., choral reading); teaching strategies; exposure to various forms of print (nonfiction, letters, poems, etc.)
    - Guided Reading and Independent Reading (40 min) - solo, buddy, or small group reading
3. Writing (35 minutes)
    - Writing Demonstrations (10 min) - composed of 3 types of activities: Write Alouds - teacher speaks aloud while demonstrating writing; Shared/Interactive Writing - students contribute to the writing process; and Mini-lessons - taught to highlight particular writing needs
    - Guided and Independent Writing - (25 min)

The results of this longitudinal study are reported in the next chapter.

5 Analysis, Results, and Discussion

The data analysis was conducted in a number of stages. First, for each grade and for both fall and spring testings, we checked for differences between the performances of boys and girls and between the performances of schools. Where differences were found, we analyzed the data further to identify precisely where the differences lay.

Second, for each grade, we looked for differences in performance between the fall and spring testings. Differences in favour of the spring testings were expected because the children were eight months older, they had an additional eight months of schooling, and the tests were the same. Therefore, improvements in the spring over the fall must be interpreted carefully. In order to assist in the interpretation, we compared the performance of the children in the fall and spring to a normative sample of their peers tested at the same time.

Third, against the spring testing in one year, we compared the fall testing of the subsequent year. In this case, although the children were four months older and tested using the same measures, only two additional months of schooling and a two-month vacation intervened. This comparison allows us to test for diminished or improved performance over the summer period.

Finally, we analyzed the data to test for differences between the students taught the M.A.P. program and students taught other reading programs in comparable schools, who served as a control group.

The results are shown by grade.

GRADE 1

Grade 1 data consisted of raw scores for the fall and spring testing on three measures:

1. Scores on the Group Reading Assessment and Diagnostic Evaluation (GRA+DE), Level K, consisting of five diagnostic test scores and their associated subtests, and scores on three other tests of GRA+DE, Level 1;
2. Test of Written Spelling (TWS); and
3. Oral and Written Language Scales (OWLS).

Tables 5.1 to 5.3 provide means, standard deviations, and numbers of children by male, female total group, and by school on these measures.

Table 5.1 Descriptive Statistics by Gender and Total Group for the Fall and Spring Testing of Grade 1 Students

	Fall Testing									Spring Testing								
	Female			Male			Total			Female			Male			Total		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
I. Group Reading Assessment and Diagnostic Evaluation																		
1. Phonological Awareness[K]	129	18.4	4.7	115	16.8	6.5	244	17.6	5.7	124	22.4	4.2	108	21.5	4.7	232	22	4.5
Sound Matching[K]	131	7.4	2.9	116	6.7	3.4	247	7.1	3.2	124	9.9	2.5	110	9.3	2.8	234	9.6	2.6
Rhyming[K]	129	10.9	2.5	115	10.1	3.7	244	10.5	3.2	124	12.5	2.2	108	12.1	2.5	232	12.3	2.4
2. Early Literacy Skills[K]	126	22.3	2.3	115	21	4.1	241	21.7	3.3	123	23.7	0.8	108	23.1	1.8	231	23.4	1.4
Print Awareness[K]	131	3.2	1.0	115	2.9	1.1	246	3	1.0	124	3.8	0.4	110	3.7	0.6	234	3.8	0.5
Letter Recognition[K]	131	10.4	1.4	119	10	2.0	250	10.2	1.7	126	10.8	0.8	110	10.5	1.2	236	10.7	1.1
Same and Different Words[K]	129	8.6	0.9	116	8.1	1.6	245	8.4	1.3	123	8.9	0.4	108	8.9	0.5	231	8.9	0.5
3. Phoneme-Grapheme Correspondence[K]	135	13	4.0	117	12	4.6	252	12.6	4.3	128	15.3	1.8	109	14.6	2.9	237	15	2.4
4. Listening Comprehension[K]	131	16.4	2.9	119	15.1	3.8	250	15.7	3.4	126	17.6	0.8	110	17.1	1.6	236	17.4	1.3
Total Score (Sum of 1-4)	126	70.3	11.3	111	65.2	16.1	237	67.9	14.0	122	79.2	6.2	107	76.5	9.3	229	77.9	7.9
5. Word Reading[K]	134	6.9	2.9	117	6.3	2.9	251	6.6	2.9	128	9.2	1.7	111	8.4	2.7	239	8.9	2.2
6. Word Meaning[1]	116	15.6	6.9	93	14.2	8.1	209	14.9	7.4	122	22.7	6.2	106	22	7.4	228	22.4	6.8
7. Sentence Comprehension[1]	113	3.2	4.7	98	3.3	5.6	211	3.2	5.1	120	11.5	6.4	109	11.5	6.9	229	11.5	6.6
8. Listening Comprehension[1]	127	14.9	2.4	105	14.4	2.9	232	14.7	2.6	126	15.6	2.3	110	15.3	2.9	236	15.4	2.6
II. Test of Written Spelling	123	3	3.1	106	2.9	3.4	229	3	3.2	112	8.9	4.7	99	8	5.7	211	8.5	5.2
III. Oral and Written Language Scales	124	16.3	5.4	105	15.5	6.0	229	15.9	5.7	122	36.5	13.0	106	32.7	14.5	228	34.7	13.8

K= Kindergarten/Grade 1 Transition Measures
1 = Grade 1 Measures

Table 5.2 Descriptive Statistics by School for the Fall Testing of Grade 1 Students

	School 1			School 2			School 3			School 4			School 5			School 6			School 7		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
I. Group Reading Assessment and Diagnostic Evaluation																					
1. Phonological Awareness[K]	49	13.4	4.9	12	19.0	4.8	33	18.2	5.2	30	17.5	5.2	38	14.7	5.6	42	19.3	4.0	40	22.3	2.7
Sound Matching[K]	51	4.6	2.4	12	7.6	2.8	33	7.4	2.8	30	7.0	3.1	39	5.2	2.6	42	7.9	2.5	40	10.6	1.4
Rhyming[K]	49	8.7	3.3	12	11.4	2.6	33	10.8	3.3	30	10.5	2.7	38	9.4	3.7	42	11.3	2.4	40	12.6	1.9
2. Early Literacy Skills[K]	49	19.0	4.2	12	22.8	1.7	33	22.3	2.5	30	22.0	2.5	37	20.4	4.2	40	22.9	1.5	40	23.7	0.5
Print Awareness[K]	51	2.2	1.0	12	3.5	0.7	33	3.1	1.1	30	3.0	1.0	38	2.8	1.0	42	3.4	0.9	40	3.8	0.4
Letter Recognition[K]	56	9.1	2.4	12	10.7	0.6	33	10.6	1.2	31	10.3	2.1	38	10.0	1.9	40.0	10.6	0.7	40	10.9	0.2
Same and Different Words[K]	50	7.7	1.6	12	8.6	0.9	33	8.7	0.7	30	8.7	0.7	38	7.6	2.2	42	8.9	0.4	40	9.0	0.0
3. Phoneme-Grapheme Correspondence[K]	58	8.5	4.6	12	14.7	2.0	33	13.2	3.7	31	12.8	3.9	38	10.9	4.0	41	15.5	1.0	39	15.9	0.3
4. Listening Comprehension[K]	56	14.0	3.4	12	16.5	2.5	33	15.4	3.9	31	15.9	4.0	38	15.0	4.4	40	17	1.4	40	17.6	1.1
Total Score (Sum of 1-4)	47	55.3	13.4	12	73.0	9.0	33	69.1	12.3	30	68.2	12.4	37	61.0	15.6	39	74.5	5.8	39	80.4	3.2
5. Word Reading[K]	55	4.0	2.0	12	8.8	2.0	33	6.5	2.8	31	5.4	3.0	40	6.7	2.6	41	8.1	2.1	39	9.2	1.2
6. Word Meaning[1]	20	10.8	6.1	13	14.9	7.5	33	13.4	7.3	32	11.6	5.8	32	11.5	6.9	40	17.2	6.4	39	21.6	5.6
7. Sentence Comprehension[1]	20	1.6	2.1	13	2.6	4.1	33	2.1	4.6	32	0.7	1.4	39	0.9	3.1	35	3.9	5.3	39	9.0	6.0
8. Listening Comprehension[1]	34	13.8	3.1	13	15.3	1.4	33	14.2	3.7	32	15.4	1.9	40	13.7	3.5	41	15.1	1.3	39	15.5	1.2
II. Test of Written Spelling	33	1.0	1.7	13	1.5	1.3	33	1.7	2.3	33	1.3	2.1	38	2.3	2.5	40	4.5	3.3	39	6.7	2.9
III. Oral and Written Language Skills	33	12.2	3.8	13	15.5	6.2	33	14.4	3.3	32	12.9	4.0	39	13.6	3.9	40	20.2	6.9	39	20.7	3.3

K= Kindergarten/Grade 1 Transition Measures
1 = Grade 1 Measures

Table 5.3 Descriptive Statistics by School for the Spring Testing of Grade 1 Students

	School 1			School 2			School 3			School 4			School 5			School 6			School 7		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
I. Group Reading Assessment and Diagnostic Evaluation																					
1. Phonological Awareness[K]	44	17.1	4.8	11	22.6	2.5	33	22.4	4.0	27	21.4	4.8	41	22.5	4.1	38	24.2	2.0	38	24.7	2.0
Sound Matching[K]	46	6.7	3.0	11	9.5	2.5	33	10.0	2.1	27	9.8	2.2	41	9.8	2.3	38	11.0	1.0	38	11.3	1.2
Rhyming[K]	44	10.5	2.7	11	13.1	0.9	33	12.4	2.4	27	11.6	3.1	41	12.7	2.1	38	13.2	1.4	38	13.4	1.2
2. Early Literacy Skills[K]	43	22.1	2.3	11	23.6	1.2	33	23.6	0.9	27	23.8	0.4	41	23.8	0.5	38	23.4	1.3	38	23.9	0.2
Print Awareness[K]	46	3.4	0.8	11	4.0	0.0	33	3.9	0.3	27	4.0	0.0	41	3.9	0.3	38	3.8	0.4	38	4.0	0.2
Letter Recognition[K]	46	9.9	1.9	12	11.0	0.0	33	10.8	0.7	28	10.9	0.3	41	11.0	0.2	38	10.6	1.1	38	11.0	0.2
Same and Different Words[K]	43	8.7	0.8	11	8.6	1.2	33	9.0	0.0	27	8.9	0.3	41	8.9	0.3	38	9.0	0.0	38	9.0	0.0
3. Phoneme-Grapheme Correspondence[K]	47	12.6	4.2	12	15.3	0.7	33	15.6	1.4	28	15.5	1.5	41	15.7	0.9	38	15.6	0.7	38	15.9	0.3
4. Listening Comprehension[K]	46	16.7	2.0	12	17.7	0.5	33	17.5	1.1	28	17.2	1.2	41	17.4	1.1	38	17.7	0.7	38	17.8	0.4
Total Score (Sum of 1-4)	41	68.3	11.1	11	79.5	2.4	33	79.1	6.4	27	77.9	6.4	41	79.4	5.6	38	80.9	3.4	38	82.3	2.2
5. Word Reading[K]	47	6.1	3.3	12	9.5	0.8	33	9.5	1.4	28	8.9	2.0	41	9.6	0.8	40	9.6	0.8	38	9.9	0.5
6. Word Meaning[1]	40	14.2	9.0	11	25.7	1.3	33	25.1	3.2	27	21.0	6.8	41	23.7	4.4	39	25.5	1.4	37	24.0	5.7
7. Sentence Comprehension[1]	41	4.5	4.9	11	15.0	4.2	33	15.1	4.8	27	7.7	6.4	41	12.7	5.9	40	13.0	5.3	36	15.1	5.3
8. Listening Comprehension[1]	47	13.0	4.6	11	16.4	0.9	33	15.7	1.6	27	16.0	1.0	41	16.0	1.3	40	16.1	1.1	37	16.2	1.1
II. Test of Written Spelling	39	3.2	3.7	12	8.9	3.0	33	10.8	5.4	26	7.0	4.1	41	8.5	4.3	23	8.6	2.0	37	13.0	4.3
III. Oral and Written Language Scales	39	18.7	11.4	12	31.7	7.9	33	40.5	10.4	26	21.5	4.9	41	34.8	10.0	39	44.5	7.6	38	46.0	8.8

K= Kindergarten/Grade 1 Transition Measures
1 = Grade 1 Measures

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 two factor Multivariate Analyses of Variance (MANOVA). The first factor was Gender with two levels, male and female. The second factor was School with seven levels, corresponding to the seven schools involved in the study. The multivariate F-test for Gender was not significant for either the fall or the spring testing. Therefore, we must conclude that the girls and boys performed equally. The performances of children from the seven schools generally were significantly different from one another across the ten measures [F (60, 980) = 4.41, p < .05] for the fall testing; F (60,998) = 6.10, p < .05, for the spring testing]. There also were significant differences on each of the ten tests taken separately (p < .05). In order to determine which pairwise differences between schools were significant on these ten variables, the Dunnett C method of multiple comparisons was chosen. The overall results support the conclusion that School 7 outperformed more schools on more measures than any other school; School 6 was second; and School 1 fared the worst.

Were there reliable or statistically significant differences in the scores from the two testing sessions in Grade 1? To address this question the data from the two testing sessions on the 10 tests, not including the GRA+DE total score, were analyzed in a Repeated Measures MANOVA, with two between factors, Gender and School. The number of cases for which there was complete data is much reduced from the total number of children involved in the study. There were complete data for 148 children. The overall multivariate result indicated that the scores in the spring session on the 10 tests were statistically better than in the fall session [F (10, 125) = 150.01, p < .05]. On each of the 10 tests, the individual result was also significant, indicating that each spring score mean was significantly better than its counterpart in the fall. There were no overall gender differences in the average scores from the two testing sessions. We can conclude that boys and girls performed on the average similarly, considering the scores from both the testing sessions. The differences between the profiles of combined means of schools in the multivariate repeated measures analysis were significant, [F (60, 740) = 4.26, p < .05]. The multiple comparisons for the means of schools followed generally the pattern of results that was found for each testing session.

In order to determine whether the improvements between fall and spring were what would normally be expected, or less or more than expected, we turned to norms for comparable students tested at the same times. Norms for the individual diagnostic tests of GRA+DE are available only as stanines. Stanines are single digit standard scores. The stanine scale divides the distribution of scores into nine parts, from 1 to 9. The stanine 5 is the middle or the average score on the stanine scale, and covers 20% of the examinees in a normally distributed set of scores, ranging from a percentile rank of 41 to 60. A percentile rank indicates an examinee's relative standing in a group in terms of the percentage of group members from a normative sample scoring at or below the examinee's

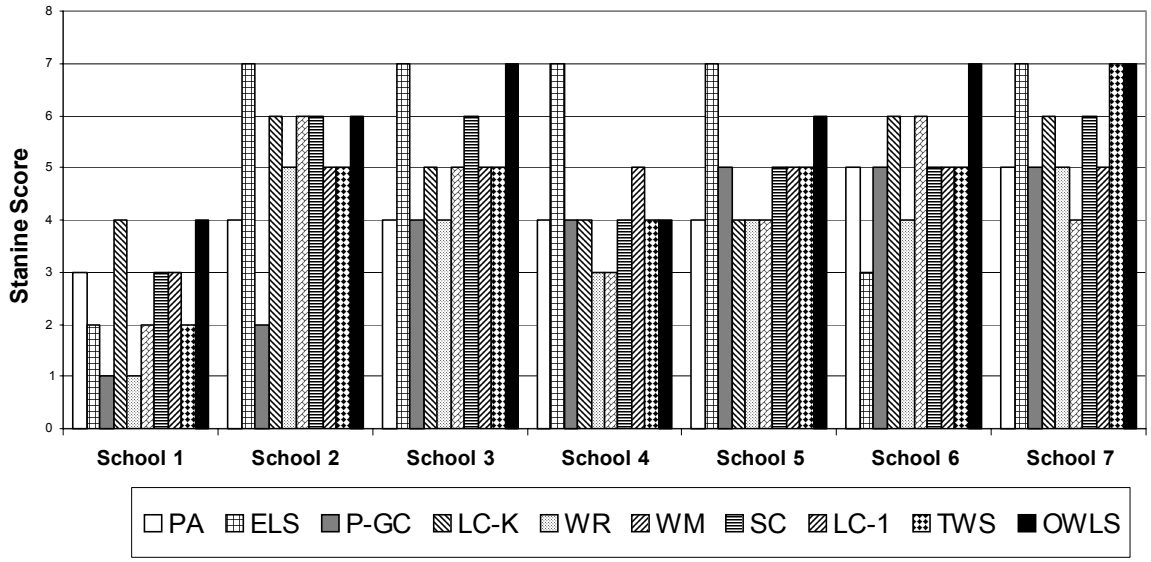
raw score. Figure 5.1 provides a correspondence between stanines and percentiles.

Figure 5.1 Correspondence between Stanine Scores and Percentile Ranks

Stanine	1	2	3	4	5	6	7	8	9
% in band	4	7	12	17	20	17	12	7	4
Percentile	4	11	23	40	60	77	89	96	100

Figure 5.2 presents the stanine scores for the spring testing by school for ten of the measures. We note that School 1 was below the average stanine score of 5 on all ten measures and School 4 was below on nine of the measures. The best performing schools, 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 average, to slightly above average.

Figure 5.2 Normative Scores for Grade 1 Spring Testing by School



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 5.3 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 5.3 Stanine Difference from Fall to Spring Testing for Grade 1 by School

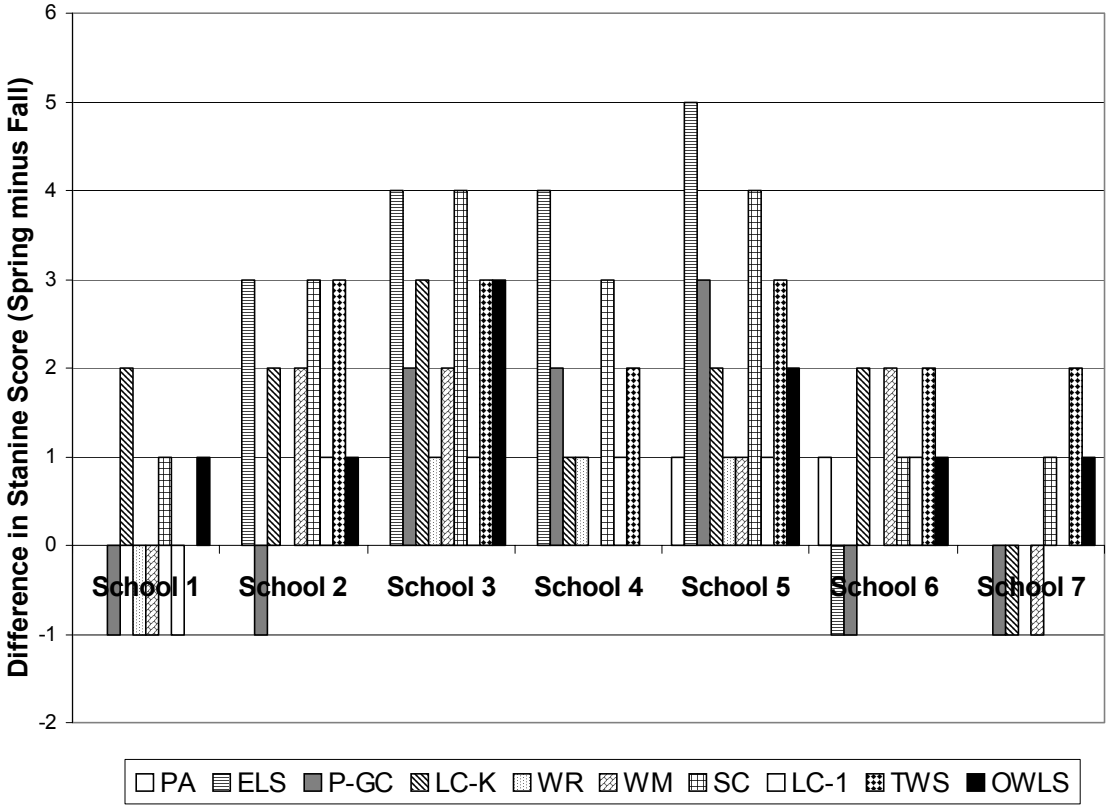
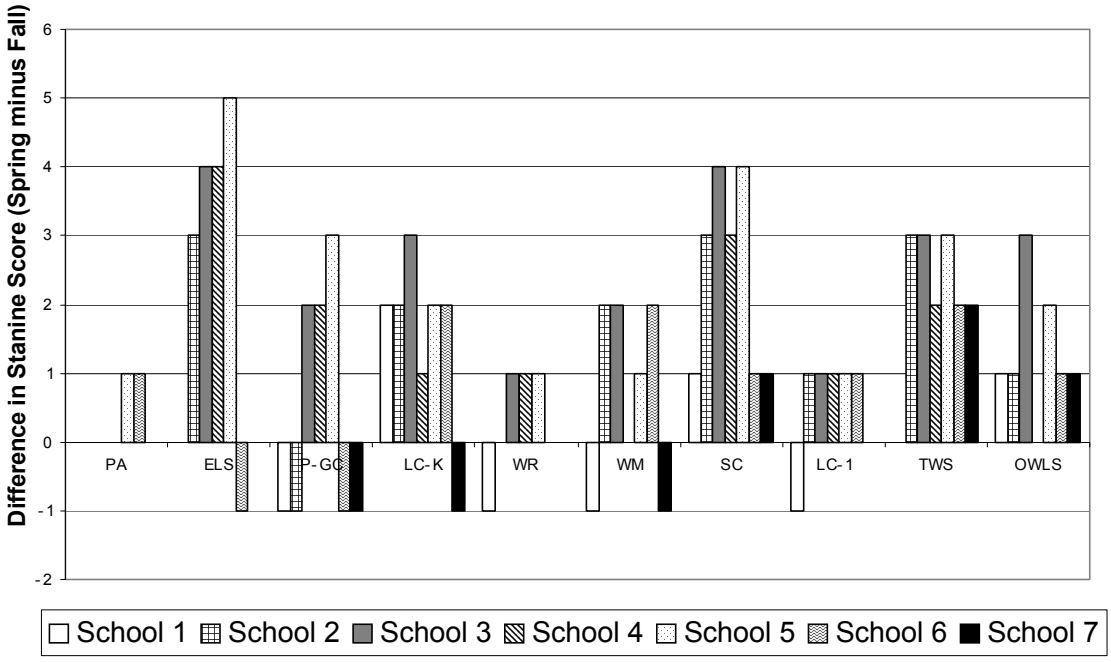


Figure 5.4 displays the same information, but grouped according to the ten measures. All seven schools showed improved normative standing on Sentence Comprehension for spring over fall. The Sentence Comprehension measure capitalizes upon students' ability to determine, based upon the context, the best word to complete a sentence. The measures on which fewest schools showed improved normative standing were Phonological Awareness, Phoneme-Grapheme Correspondence, and Word Reading. Phoneme-Grapheme Correspondence was also the measure on which the most schools showed a decline in normative standing — four schools. Given the emphasis on phonological awareness and phoneme-grapheme correspondence in the M.A.P. program, and given the research evidence discussed in Chapter 2 showing that phonological awareness instruction has limited impact on reading, these results are difficult to interpret. These results do raise the question about program implementation fidelity, and whether schools placed less emphasis on phonological awareness than called for by the M.A.P. program and more on other aspects of reading.

Figure 5.4 Stanine Difference from Fall to Spring Testing for Grade 1 by Test



GRADE 2

We obtained scores for Grade 2 children on GRA+DE Level 2: Vocabulary Composite (Word Reading and Word Meaning), Comprehension Composite (Sentence Comprehension and Passage Comprehension); and Listening Comprehension. We also obtained Test of Written Spelling and Oral and Written Language Scales results. Tables 5.4 to 5.6 provide descriptive statistics for all the subtest, composite, and total test scores for Grade 2 children by gender, total group, and school.

In order to test the hypothesis of equality of means on the five variables (Vocabulary Composite, Comprehension Composite, Listening Comprehension, TWS, and OWLS) two factor MANOVAs were conducted. The multivariate test for Gender was significant for the fall testing [F (5, 138) = 2.71, $p < .05$], indicating that the performance profiles were significantly different for girls and boys. However, none of the corresponding univariate tests of the five variables taken individually was significant, so we are unable to point to individual measures on which girls outperformed boys or boys outperformed girls. The MANOVA results indicated that there were no significant differences between the profiles of means of females and males on each of the five GRA+DE dependent variables for the spring testing.

Table 5.4 Descriptive Statistics by Gender and Total for the Fall and Spring Testing of Grade 2 Students

	Fall Testing									Spring Testing								
	Female			Male			Total			Female			Male			Total		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
I. Group Reading Assessment and Diagnostic Evaluation Level 2																		
1. Vocabulary Composite	94	44.5	13.0	76	46.2	11.8	170	45.3	12.5	96	50.6	7.3	71	50.6	8.7	167	50.6	7.9
Word Reading	95	22.4	6.3	79	22.3	6.9	174	22.4	6.6	97	25.6	3.6	72	25.6	4.2	169	25.6	3.9
Word Meaning	99	22.1	7.0	76	23.2	6.1	175	22.6	6.6	97	25	3.9	74	24.6	5.4	171	24.8	4.6
2. Comprehension Composite	95	25.0	13.2	76	25.1	12.5	171	25.1	12.9	93	34.3	10.4	74	32.8	10.8	167	33.6	10.6
Sentence Comprehension	97	11.4	6.1	76	12.2	5.6	173	11.8	5.9	97	15.4	4.6	75	15.0	4.9	172	15.2	4.7
Passage Comprehension	98	13.5	7.8	77	12.7	7.6	175	13.1	7.7	94	18.9	6.3	74	17.8	6.3	168	18.4	6.3
Total Score (Sum of 1and 2)	92	69.7	24.0	75	71.9	21.4	167	70.7	22.8	92	85.1	17.0	70	83.9	18.1	162	84.6	17.5
3. Listening Comprehension	94	14.9	2.1	79	14.7	1.8	173	14.8	2.0	96	15.5	1.6	70	15.0	2.2	166	15.3	1.9
II. Test of Written Spelling	100	8.9	5.4	74	10.7	6.2	174	10.0	5.8	99	15.5	6.8	75	16.6	7.8	174	16	7.2
III. Oral and Written Language Scales	99	28.5	18.0	78	26.0	18.8	177	27.4	18.4	95	41.5	14.5	76	37.0	17.8	171	39.5	16.2

Table 5.5 Descriptive Statistics by School for the Fall Testing of Grade 2 Students

	School 1			School 2			School 3			School 4			School 5			School 6			School 7		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
I. Group Reading Assessment and Diagnostic Evaluation Level 2																					
1. Vocabulary Composite	24	25.2	13.7	9	52.0	3.2	18	51.6	4.4	22	40.5	13.3	33	47.1	9.1	35	49.0	6.2	29	53.0	3.2
Word Reading	27	12.0	6.2	9	25.6	3.0	18	25.4	3.1	22	20.9	6.4	33	23.0	5.2	35	24.4	3.9	30	26.9	1.7
Word Meaning	28	13.5	8.4	9	26.4	0.7	18	26.1	1.5	22	19.6	7.5	33	24.1	4.5	36	24.6	2.8	29	26.2	1.6
2. Comprehension Composite	25	12.2	4.3	9	33.0	7.9	18	31.2	10.9	22	17.1	12.1	33	22.6	11.4	35	27.3	11.6	29	36.0	10.0
Sentence Comprehension	26	5.2	3.1	9	14.8	3.9	18	15.2	4.6	22	7.9	5.0	33	11.9	5.4	35	12.7	5.3	30	16.0	4.2
Passage Comprehension	26	7.1	3.3	9	18.2	4.5	18	15.9	7.2	23	9.1	7.6	33	10.7	7.2	36	14.0	6.9	30	19.8	6.2
Total Score (Sum of 1and 2)	22	37.3	14.1	9	85.0	11.0	18	82.7	14.7	21	59.5	21.6	33	69.7	19.0	35	76.0	16.6	29	87.1	15.8
3. Listening Comprehension	26	12.2	1.8	9	15.0	1.0	18	15.6	1.0	22	14.8	1.5	33	14.3	2.4	35	15.7	1.3	30	15.9	0.9
II. Test of Written Spelling	29	4.1	3.5	8	7.9	4.4	17	10.2	5.4	22	8.2	4.8	34	8.9	4.3	36	11.3	5.6	28	15.6	5.1
III. Oral and Written Language Scales	33	1.7	11.7	8	21.5	13.8	17	32.3	10.4	23	20.9	16.5	33	33.7	10.8	36	36.6	11.4	27	43.3	10.8

Table 5.6 Descriptive Statistics by School for the Spring Testing of Grade 2 Students

	School 1			School 2			School 3			School 4			School 5			School 6			School 7		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
I. Group Reading Assessment and Diagnostic Evaluation Level 2																					
1. Vocabulary Composite	26	40.5	13.0	10	53.5	1.4	18	54.3	1.2	17	45.9	10.2	34	52.8	2.9	33	52.9	2.0	29	54.0	1.5
Word Reading	27	20.8	6.3	10	27.1	1.1	18	27.4	1.0	18	23.4	4.5	34	26.7	1.5	33	26.7	1.5	29	27.2	1.2
Word Meaning	28	19.5	7.3	10	26.4	0.7	18	26.9	0.5	19	22.3	6.5	34	26.0	1.7	33	26.2	0.8	29	26.7	0.7
2. Comprehension Composite	28	19.7	8.9	10	39.2	6.5	18	39.5	5.1	18	26.0	12.5	32	35.4	7.7	32	36.0	6.7	29	41.7	3.0
Sentence Comprehension	28	8.9	4.5	10	17.9	2.5	18	17.7	1.7	20	11.7	5.8	34	16.1	3.0	33	17.0	2.5	29	18.1	1.3
Passage Comprehension	29	11.0	5.1	10	21.3	4.2	18	21.8	3.8	18	14.7	7.0	32	19.5	5.1	32	18.8	5	29	23.6	2.2
Total Score (Sum of 1and 2)	26	60.2	20.9	10	92.7	7.6	18	93.8	5.9	15	71.9	22.3	32	88.1	10.0	32	89.1	8.1	29	95.7	4.2
3. Listening Comprehension	23	13.4	2.1	10	16.6	0.5	18	15.7	1.2	20	15.2	1.5	33	14.9	2.6	33	15.6	1.2	29	16.0	1.2
II. Test of Written Spelling	28	7.2	5.2	11	15.4	5.2	18	16.8	5.4	22	13.5	7.4	33	17.2	5.6	32	18.9	5.6	30	21.2	5.6
III. Oral and Written Language Scales	26	16.0	13.6	11	42.5	11.8	18	46.9	13.5	22	31.8	16.7	33	41.1	9.9	31	45.9	8.7	30	51.6	7.7

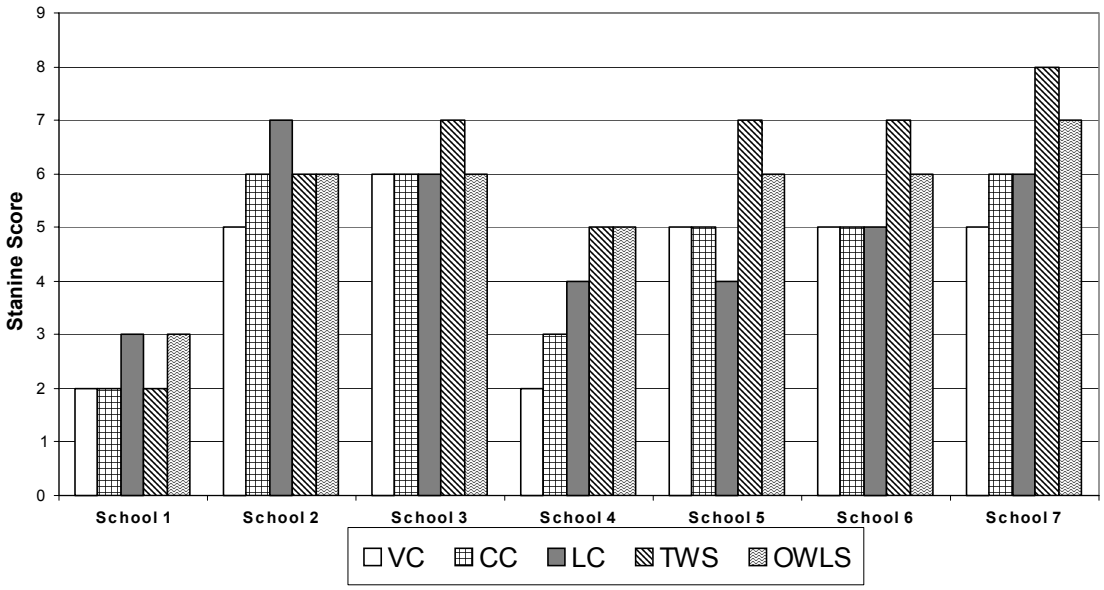
The multivariate School effect on the dependent variables Vocabulary Composite, Comprehension Composite, Listening Comprehension, TWS, and OWLS was significant [F (30, 682) = 10.69, $p < .05$ for fall; F (30, 652) = 6.06, $p < .05$ for spring]. The univariate School effects were also significant for the five dependent variables in both fall and spring ($p < .05$). The Gender by School interaction was not significant in the case of either testing session. Therefore, School differences were not confounded by the gender factor; the overall performance of schools was the same for both girls and boys. The post hoc analysis was conducted using the Scheffé method of multiple comparisons in order to see wherein the school differences lay. The analyses showed School 7 was the highest performing school and School 1 was the lowest performing. School 2 was the second highest performing and School 4 was the second lowest performing overall.

Were there statistically significant differences in the raw scores from the two testing sessions in Grade 2, fall and spring? The data from the two testing sessions were analyzed in a Repeated Measures MANOVA, with two between factors, Gender and School, and the following dependent measures: Vocabulary Composite, Comprehension Composite, Listening Comprehension, Test of Written Spelling, and Oral and Written Language Scales.

Because of the listwise deletion of missing cases, the number of cases in this analysis is reduced to 144. The overall multivariate result indicated that the scores in the spring session on the 5 measures were statistically better than in the fall session [F (5, 1106) = 102.32, $p < .05$]. On each of the 5 dependent variables, the individual result was also significant, demonstrating that each spring score mean was significantly better than its counterpart in the fall. There were no overall gender differences in the average scores from the two testing sessions, which is consistent with the results for Gender for each session. The differences between the profiles of combined means of schools in the multivariate repeated measures analysis were significant, [F (30, 522) = 6.33, $p < .05$], similar to the results for the fall and spring sessions taken separately. The multiple comparisons for the means of schools followed generally the pattern of results that was found for each testing session with School 7 performing the best and School 1 the worst.

Looking to Figure 5.5, we see that as in Grade 1, School 1 remains below the average stanine of 5 on all measures. Indeed, the students performed more poorly in Grade 2 compared to the normative sample. School 4 also fared worse than in Grade 1 compared to the normative sample with no performance above the average stanine. Schools 2 and 7 performed highly as in Grade 1, with scores above average on four out of five measures. School 3 scored above average on all measures, showing an improvement in standing over Grade 1. The students in School 6 showed a performance in the middle of the group as they had in Grade 1. Except for School 7, all the schools displayed a smaller difference between the measures on which they received their lowest normative rankings and the ones on which they achieved their highest. This result indicates that students' abilities across the various aspects of reading and writing had become more equal.

Figure 5.5 Normative Scores for Grade 2 Spring Testing by School



According to Figure 5.6, the schools showing improvement in normative ranking on the highest number of measures were Schools 2, 3, and 6, with improvements on 4 measures each. Only three schools showed lower rankings compared to the fall testing, and then on only one measure each. Figure 5.7 shows that the greatest improvements were on TWS rankings, with every school showing improvement except School 1. Two schools showed declines in rankings on the Vocabulary Composite (Schools 4 and 7) and one school (School 1) showed a decline on the Comprehension Composite.

Figure 5.6 Stanine Difference from Fall to Spring Testing for Grade 2 by School

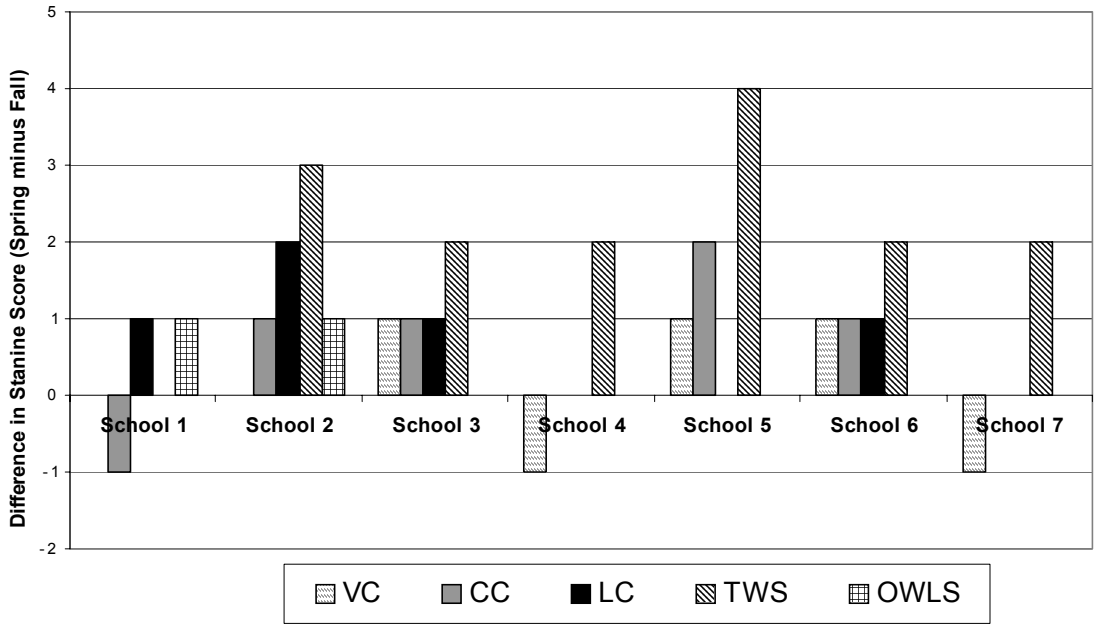
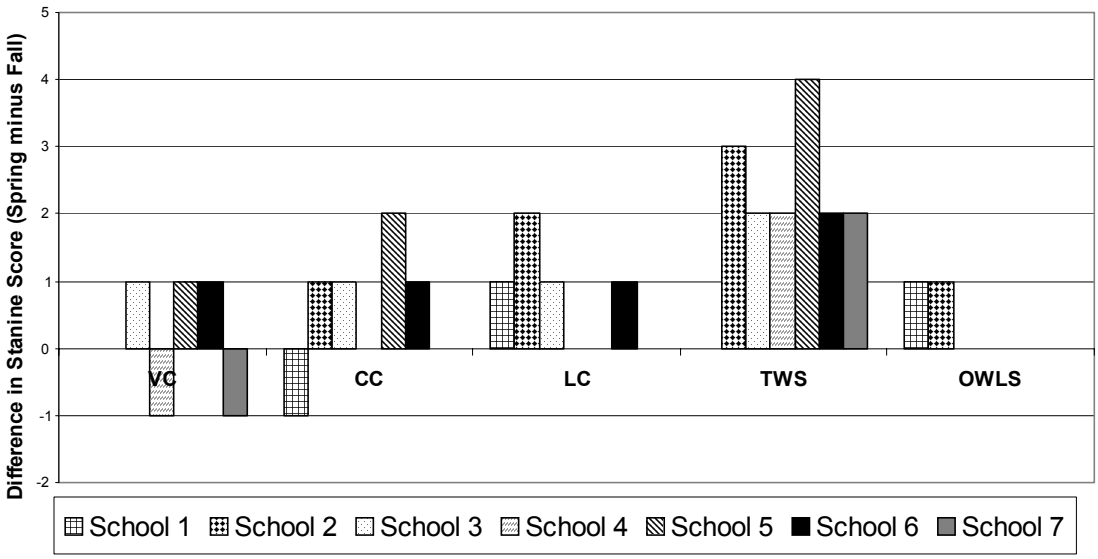


Figure 5.7 Stanine Difference from Fall to Spring Testing for Grade 2 by Test



GRADE 3

The source of data for Grade 3 was GRA+DE Level 3, TWS, and OWLS tests. Descriptive statistics for all the subtest, composite, and total test scores for Grade 3 children by gender, total group, and school are given in Tables 5.7 to 5.9.

In order to test the hypothesis of equality of the profiles of means on five variables (Vocabulary Composite, Comprehension Composite, Listening Comprehension, TWS, and OWLS) for Gender and Schools, data were analyzed in two factor MANOVAs. The multivariate test for Gender was not significant for the fall testing, indicating that the two profiles were not significantly different. The MANOVA results for the spring testing indicated that there were significant differences between the profiles of means of females and males [F (5, 110) = 4.54, $p < .05$] with the girls scoring significantly higher on the Vocabulary Composite and on OWLS ($p < .05$). The multivariate School effect was significant [F (30, 497) = 6.08, $p < .05$ for fall; F (30, 542) = 8.46, $p < .05$ for spring]. The univariate School effects also were significant for the five dependent variables ($p < .05$) for both testing sessions. The Gender by School interaction was not significant in either testing, so School differences are not confounded by the Gender factor. The results of the post hoc analysis for the School factor showed that School 7 was the highest performing school and School 1 was the lowest performing. School 6 was the second highest performing and School 4 was the second lowest performing overall.

Table 5.7 Descriptive Statistics by Gender and Total Group for the Fall and Spring Testing of Grade 3 Students

	Fall Testing									Spring Testing								
	Female			Male			Total			Female			Male			Total		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
I. Group Reading Assessment and Diagnostic Evaluation Level 3																		
1. Vocabulary Composite	75	47.3	11.2	52	48.2	10.6	127	47.7	10.9	77	51.9	7.9	55	50.1	10.4	132	51.1	9.0
Word Reading	76	27.4	3.8	53	27.4	4.4	129	27.4	4.1	77	28.7	1.9	55	28.0	3.9	132	28.4	2.9
Vocabulary	78	19.9	8.0	52	20.7	7.3	130	20.2	7.7	77	23.2	6.4	55	22.1	7.0	132	22.7	6.7
2. Comprehension Composite	75	31.0	11.3	51	31.5	10.5	126	31.2	10.9	75	36.1	8.4	55	34.4	9.2	130	35.4	8.7
Sentence Comprehension	76	14.7	5.1	53	15.0	4.5	129	14.8	4.9	77	16.6	3.9	55	16.6	3.6	132	16.6	3.8
Passage Comprehension	78	16.2	6.8	51	16.4	6.4	129	16.3	6.6	76	19.4	5.6	55	17.9	6.4	131	18.8	5.9
Total Score (Sum of 1and 2)	75	78.3	22.1	51	79.7	20.9	126	78.9	21.5	75	88.1	15.8	55	84.5	19.0	130	86.6	17.2
3. Listening Comprehension	76	15.0	1.4	53	14.7	1.8	129	14.9	1.6	77	14.2	1.6	55	14.2	1.5	132	14.2	1.6
II. Test of Written Spelling	79	15.7	7.4	53	15.9	8.3	132	15.8	7.7	78	20.2	7.1	55	21.0	9.1	133	20.5	8.0
III. Oral and Written Language Scales	78	47.3	15.3	53	41.9	17.6	131	45.1	16.4	78	50.0	13.2	55	44.9	16.3	133	47.9	14.7

Table 5.8 Descriptive Statistics by School for the Fall Testing of Grade 3 Students

	School 1			School 2			School 3			School 4			School 5			School 6			School 7		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
I. Group Reading Assessment and Diagnostic Evaluation Level 2																					
1. Vocabulary Composite	19	30.3	9.2	9	49.8	10.0	14	53.3	5.4	16	45.6	9.8	28	48.0	8.6	18	53.1	5.3	23	54.8	4.6
Word Reading	20	21.6	6.2	10	27.7	3.8	14	28.9	1.1	16	27.4	2.8	28	27.9	3.0	18	29.2	1.2	23	29.3	1.1
Vocabulary	21	8.9	3.9	9	22.2	6.4	14	24.4	4.9	16	18.2	7.8	28	20.2	6.6	19	24.2	4.9	23	25.5	3.9
2. Comprehension Composite	19	14.7	7.2	9	34.7	9.6	14	37.9	3.5	16	26.8	11.1	27	31.2	9.1	18	36.6	5.0	23	38.2	6.0
Sentence Comprehension	20	7.9	4.7	10	16.3	4.0	14	17.1	1.9	16	12.4	5.8	28	15.3	3.4	18	17.6	1.2	23	17.7	2.1
Passage Comprehension	21	7.0	3.3	9	18.4	5.8	14	20.9	3.2	16	14.4	5.8	27	15.9	6.3	19	19.1	4.3	23	20.4	4.3
Total Score (Sum of 1and 2)	19	44.9	15.8	9	84.4	19.3	14	91.2	8.0	16	72.4	20.3	27	79.2	17.4	18	89.7	9.9	23	93.0	10.4
3. Listening Comprehension	20	13.9	2.0	10	15.0	0.7	14	15.7	1.0	16	14.2	1.4	28	14.8	2.0	18	15.3	1.1	23	15.4	1.1
II. Test of Written Spelling	22	6.6	2.8	9	14.9	6.5	14	18.5	6.3	17	12.7	6.8	28	16.0	6.3	17	20.2	6.7	25	21.5	6.5
III. Oral and Written Language Scales	19	19.1	12.7	10	47.1	14.1	15	50.2	12.0	15	40.9	15.7	28	42.0	8.8	19	54.9	6.9	25	59.5	6.2

Table 5.9 Descriptive Statistics by School for the Spring Testing of Grade 3 Students

	School 1			School 2			School 3			School 4			School 5			School 6			School 7		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
I. Group Reading Assessment and Diagnostic Evaluation Level 2																					
1. Vocabulary Composite	24	37.3	9.3	10	53.1	7.0	15	54.7	5.9	14	51.9	6.4	28	52.2	5.1	16	56.3	2.0	25	56.4	4.1
Word Reading	24	25.2	5.1	10	28.6	2.5	15	29.1	1.7	14	28.7	1.7	28	29.0	1.2	16	29.9	0.3	25	29.2	1.2
Vocabulary	24	12.1	5.0	10	24.5	4.5	15	25.7	4.7	14	23.2	4.8	28	23.2	4.5	16	26.4	2.1	25	27.2	3.0
2. Comprehension Composite	23	22.8	7.3	10	38.9	6.9	15	40.3	5.6	14	35.0	7.8	27	34.6	6.6	16	39.7	4.3	25	41.0	4.1
Sentence Comprehension	24	11.0	4.6	10	18.3	1.6	15	18.2	1.6	14	17.4	1.7	28	16.9	3.3	16	18.4	0.8	25	18.4	1.0
Passage Comprehension	23	12.0	3.5	10	20.6	5.4	15	22.1	4.3	14	17.6	6.6	27	17.1	5.7	17	22.1	3.4	25	22.5	3.5
Total Score (Sum of 1and 2)	23	59.9	15.9	10	92.0	13.6	15	95.0	11.3	14	86.9	13.4	27	86.7	10.7	16	96.0	5.3	25	97.4	7.7
3. Listening Comprehension	24	13.0	1.5	10	13.9	0.9	15	13.9	1.1	14	14.6	1.8	28	14.5	1.8	16	15.1	1.1	25	14.7	1.2
II. Test of Written Spelling	23	10.9	5.8	10	20.3	6.2	15	20.8	7.6	15	19.3	7.3	28	21.3	5.6	17	24.4	6.5	25	26.5	6.3
III. Oral and Written Language Scales	23	24.9	13.6	10	49.1	13.2	15	54.5	10.5	15	50.9	9.7	28	47.9	7.5	17	51.9	7.9	25	59.9	5.8

Were there statistically significant differences in the raw scores from the two testing sessions in Grade 3? The data from the two testing sessions on the 5 measures, not including the GRA+DE total test scores, were analyzed in a Repeated Measures MANOVA, with two between factors, Gender and School. Complete data were available for 109 children. The overall multivariate result indicated that the scores in the spring testing on the 5 measures were statistically different, not necessarily better, than in the fall testing [F (5, 91) = 53.27, $p < .05$]. On each of the 5 dependent repeated variables, the individual result was also significant, demonstrating that each spring score mean was significantly different than its counterpart in the fall. The term "different", not "better", was used above because there was a Testing Session by School interaction [F (30, 447) = 2.72, $p < .05$]. This interaction was observed for each measure except Listening Comprehension. Therefore, differences between testing sessions depend on the school involved on the four dependent variables. There was an average systematic *decline* in achievement from the fall to the spring session on Listening Comprehension. On the other hand, on the four variables with significant interaction, the changes were unsystematic, that is, not to the same degree in each school.

Figure 5.8 shows how schools rank against the normative standard. Compared to Grades 1 and 2 (Figures 5.2 and 5.5), we note less variation among the schools. 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 fallen back against the normative sample and also rank near the average.

Figure 5.9 can be compared to Figures 5.3 and 5.6. We see that in Grade 3, the schools were doing less well when compared to the normative sample in almost as many measures as they were doing better. In Grades 1 and 2, by comparison, the schools improved in the spring over the fall in the vast majority of measures. Moreover, five schools ranked lower in the spring than the fall on at least one measure, more schools than either of the previous two years.

Figure 5.10 shows that Listening Comprehension, TWS, and OWLS were the measures on which schools fell back. Most prominent were the decline by three schools on Listening Comprehension and by four schools on OWLS. These results suggest that any positive effects of the M.A.P. program had run their course by the end of Grade 3. Students were still performing at or near average on some isolated reading and writing skills. However, their weakening performance in OWLS, which required students to read instructions and write in response to them, suggests that their knowledge of isolated skills is not translating into more genuine literacy performance.

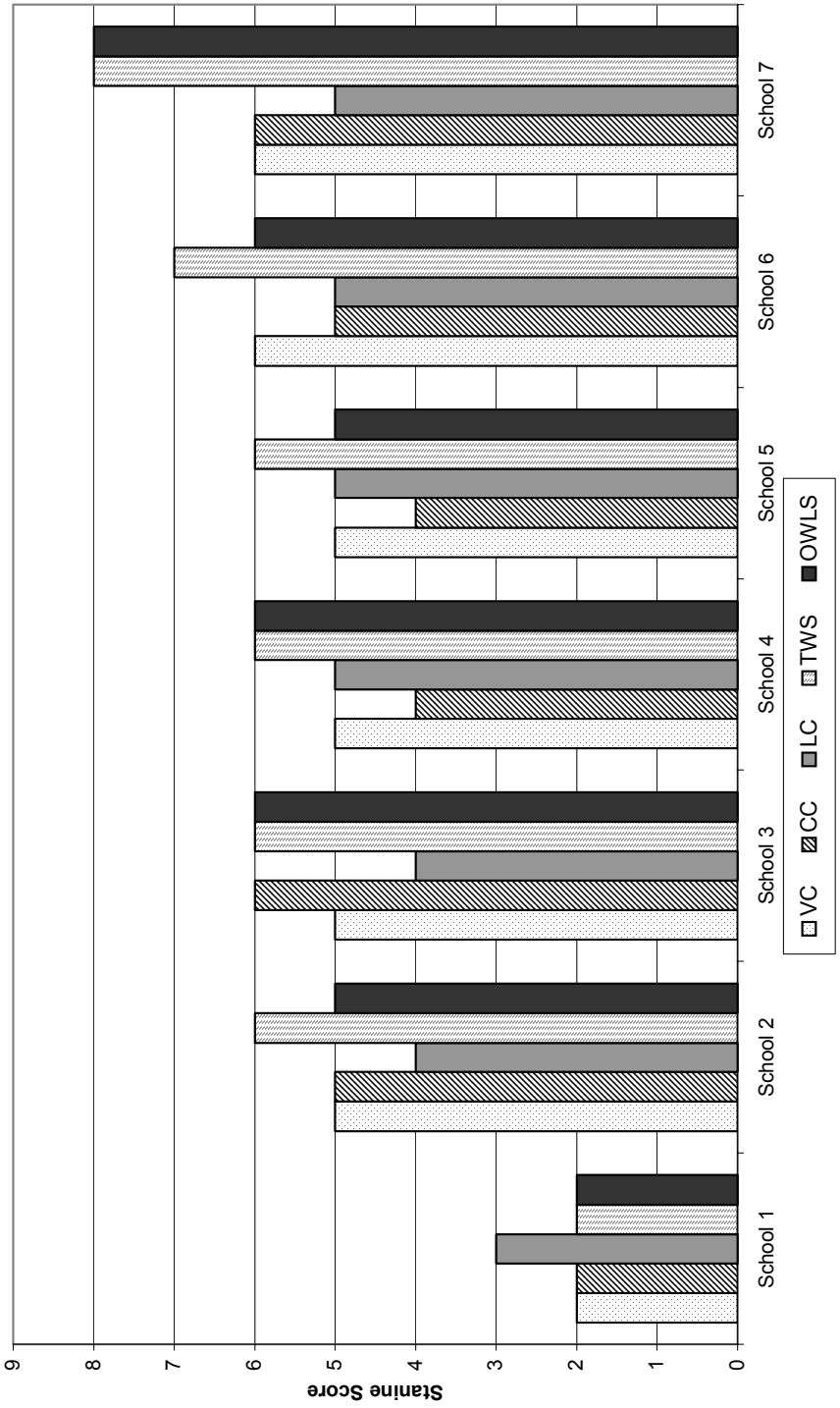


Figure 5.8 Normative Scores for Grade 3 Spring Testing by School

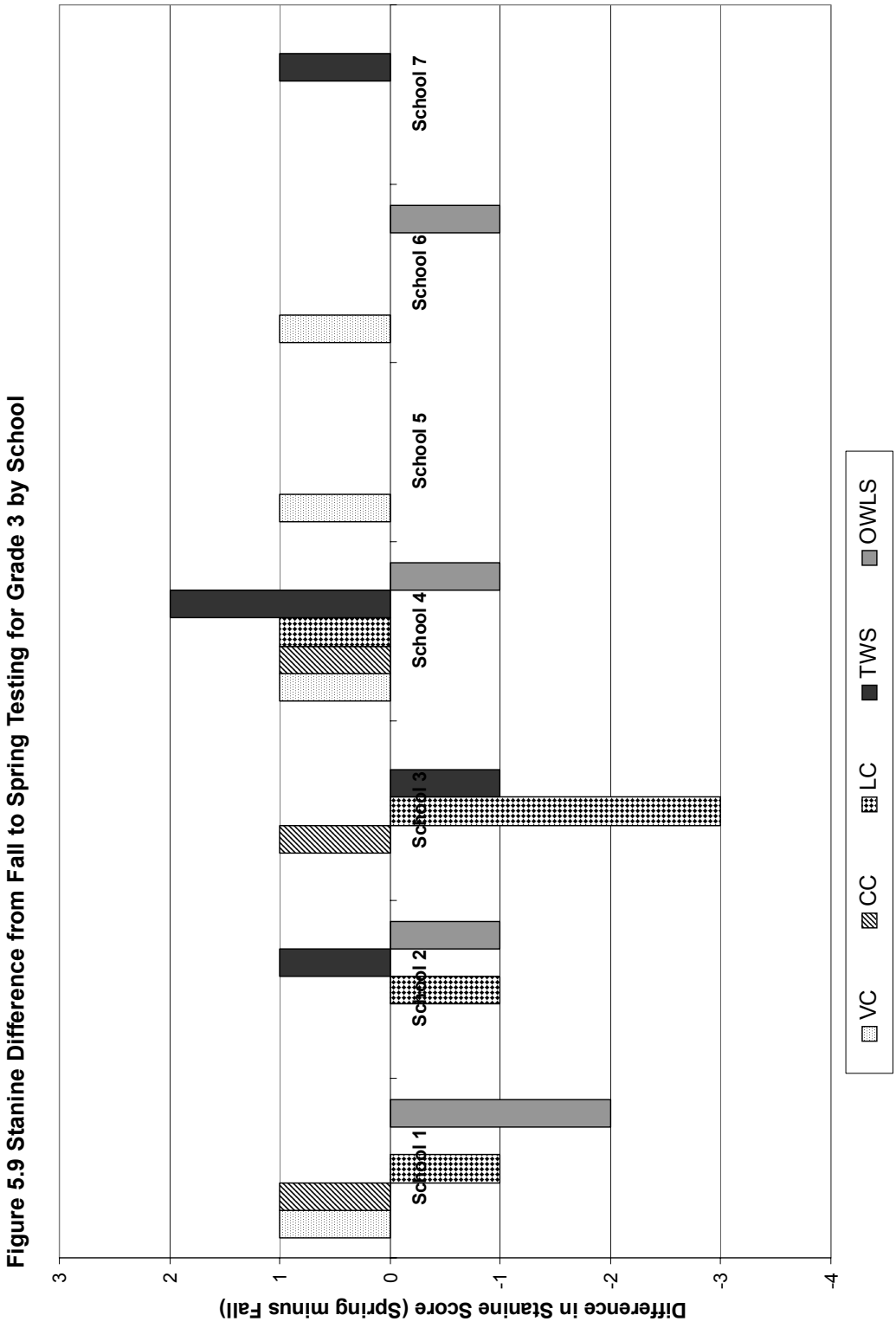


Figure 5.9 Stanine Difference from Fall to Spring Testing for Grade 3 by School

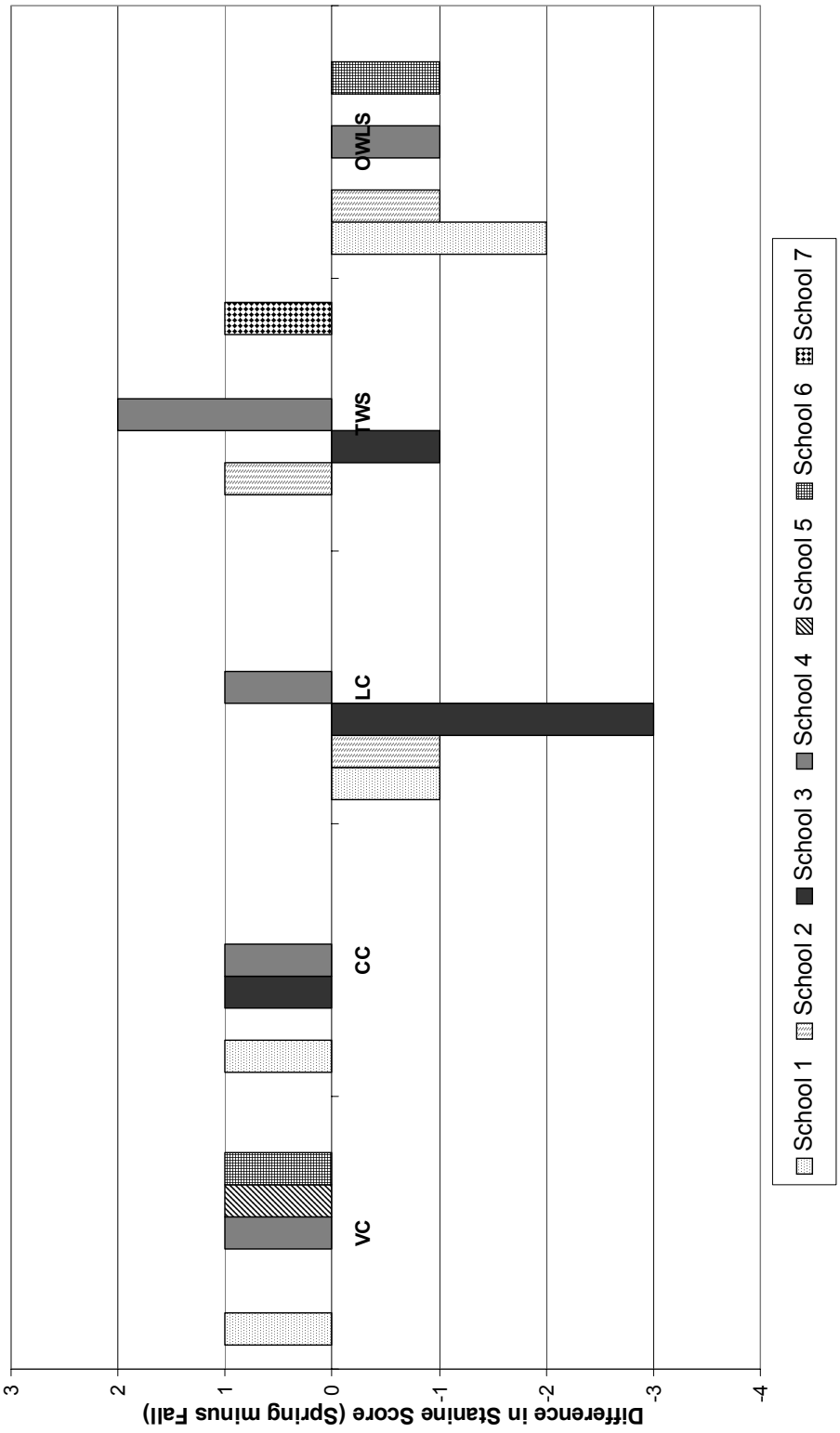


Figure 5.10 Stanine Difference from Fall to Spring Testing for Grade 3 by Test

TRANSITION BETWEEN GRADES

On the two common measures, OWLS and TWS, using repeated measures multivariate analysis of variance, we examined the stability and change in achievement of children from the spring testing in Grade 1 to the fall testing in Grade 2 and from spring Grade 2 to fall Grade 3. In this analysis the between subjects factors were Gender and School. Since Gender and School results have already been presented under each grade, these results are not presented here again. We will limit this discussion to the question of transition. What was the effect?

For the transition from Grade 1 to 2, the multivariate repeated measures result was significant [$\underline{F}$ (2, 116) = 34.65, $\underline{p} < .05$]. On examining the descriptive statistics provided in Table 5.10, children gained on both these measures from the spring to the fall testing. However, the univariate result was significant only for the OWLS [$\underline{F}$ (1, 117) = 55.81, $\underline{p} < .05$]. As all the interactions were also significant, the gain in achievement on this measure depends on both the gender and the specific school children attended. There were no differences in the means of children on TWS for the two testing sessions. However, caution is in order here. While 228 children wrote the OWLS test in the spring in Grade 1, for the transition analysis data were available on both the measures for only 131 children. Had all the children present at the beginning remained in the study until the end, the results might have been different and the differences could have gone in various directions.

For the transition from Grade 2 to 3, the multivariate repeated measures effect was significant [$\underline{F}$ (2, 104) = 11.67, $\underline{p} < .05$]. This suggests that overall there were differences in the two measures from the spring testing in Grade 2 to the fall testing in Grade 3. Again, similar to the results for the transition from Grade 1 to Grade 2, for the transition from Grade 2 to Grade 3 the univariate result was significant only for OWLS [$\underline{F}$ (1, 105) = 23.37, $\underline{p} < .05$]. As can be seen in Table 10, children in the fall testing in Grade 3 performed better than they did in the spring testing in Grade 2 on OWLS. As there were no significant interactions, the gain on this measure between the two testing occasions was systematic, that is, the gender or the school of the participating children did not differentially affect their performance. These children had similar means on TWS at the two testing occasions. As before, because there is a significant reduction in the number of children who had scores on both the measures used for these analyses caution is advised in interpreting the results.

These results are heartening. OWLS is a comprehensive measure of children's knowledge of language structure in three ways: a) their ability to use language conventions; b) their ability to use linguistic forms; and c) their ability to communicate meaningfully. The results suggest that there certainly was no loss of language ability over the summer and that there likely was some improvement. A likely explanation is that the students were engaging in language use in meaningful contexts, which enabled them to improve upon

what they already knew. Indeed the gains over the four month period that included the summer vacation were comparable to the gains over an eight month school year from the fall to the spring testings.

Table 5.10 Descriptive Statistics for the Two Measures in Transition Analyses

	Transition							
	Grade 1 to Grade 2 (n= 131)				Grade 2 to Grade 3 (n = 119)			
	Grade 1 Spring		Grade 2 Fall		Grade 2 Spring		Grade 3 Fall	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. Oral and Written Language Scales	32.8	14.3	48.4	17.2	60.9	14.7	65.3	15.0
2. Test of Written Spelling	8.3	5.5	9.8	5.9	16.8	7.1	16.7	7.4

EFFECT OF TREATMENT

Data for the participants in the Meaningful Applied Phonics (M.A.P.) program and their cohorts in the control classes 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 two between factors 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 (other authorized district reading programs with less of a phonics focus). Observed descriptive statistics for the control and treatment groups by gender are provided in Table 5.11.

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 ($p < .05$). As can be seen in Table 5.12, the predicted means from the fitted model for each of the three variables for the Control participants was significantly larger than the Treatment participants. The 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 M.A.P. program (treatment classrooms) on measures taken in Grades 2 and 3 collectively as well as individually.

The improvements shown by students grew progressively less from Grades 1 to 3. In addition, they fell back on a growing number of measures. What was not apparent from these data about the children in the M.A.P. program only is that the program was actually

detrimental to their literacy development. When the M.A.P. students are compared to their peers in other programs, we see that the M.A.P. students would have done better had they too not been in the M.A.P. program.

Table 5.11 Descriptive Statistics by Gender and Condition

	Gender									Condition					
	Female			Male			Total			Treatment			Control		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
1. Grade 1 HLAT	136	27.7	9.2	98	28.7	8.6	234	28.1	8.9	117	29.4	9	117	26.7	8.8
2. Grade 2 HLAT	136	34.3	6.3	98	34.1	6.2	234	34.3	6.2	117	34.2	6.5	117	34.3	6.0
3. Grade 3 HLAT	136	23.6	5.3	98	23.4	5.2	234	23.5	5.3	117	23.3	5.7	117	23.8	4.8
4. Grade 3 PAT	136	32.0	5.5	98	32.4	5.7	234	32.2	5.6	117	31.6	6.1	117	32.8	5.0

Table 5.12 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

THE MATTER OF SAMPLE ATTRITION

Attrition can become a problem in longitudinal studies when it lowers either internal or external validity. Attrition is always a potential matter of concern in interpreting results, so we address it here.

First, internal validity is concerned with the soundness of the inference of an effect (or not) in the particular study. In our study we concluded that the control group outperformed the treatment group (made up of those students on whom we had complete data from first to third grades). Attrition from the treatment group had absolutely no bearing upon this conclusion. The reason for this is that the control group was put together after we knew the group of treatment students on whom we had complete data and was designed to match that group of students.

Second, external validity is concerned with the soundness of inferences from the observed effects to probable effects in as yet unobserved groups (students who take the M.A.P. program subsequently in EPSB, students who take the M.A.P. program elsewhere in Alberta, students who take the M.A.P program in other provinces, etc.). Attrition from the treatment group can raise the question of whether the conclusion found in this study would apply to the entire original sample. However, this question can also be

raised about application of the conclusion to any students not actually compared in the treatment versus control analysis. This is a limitation of all educational research. There is always some group that the sample of students studied does not represent, in the very strict statistical sense that the sample is not drawn randomly from the group.

So, we are left with the question: are researchers ever justified in generalizing from any educational research? In our view, and this view is very widely accepted, if you have evidence of the ineffectiveness of a treatment within one sample, there is no scientifically or morally defensible reason for ignoring that evidence simply on the grounds that the sample is not randomly drawn from those other groups. That is a standard that is rarely reached in educational research. Rather, one acts justifiably in accord with the best available evidence.

FIDELITY OF IMPLEMENTATION AND TEACHER PREPARATION IN M.A.P.

It might be tempting also to point to poor implementation of the M.A.P. program or to poor teacher preparation for teaching it to explain the program's results in this study. We have examined these factors as described in Chapter 4 and report on them here. The results of the procedure described in Chapter 4 are given in Table 5.13.

Table 5.13 Analysis of Teacher Surveys for M.A.P. Implementation by School

School	Phonograms[a]			Spelling[a]			Number of Spelling Words per Week[b]			Reproducible Charts[b]			Homework[b]			Other[b]			Total Implementation Score
	Grade			Grade			Grade			Grade			Grade			Grade			
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
1	10	6	4	6	8	6	3	3	3	2	1	2	3	1	1	0	1	2	62
2	7	10	0	7	10	0	2	3	0	0	1	0	0	2	0	0	0	0	42
3	10	7	3	9	9	9	3	0	3	2	1	0	3	3	1	3	3	2	71
4	8	8	7	9	9	10	3	3	3	3	3	3	3	3	3	2	2	3	85
5	8	8	9	9	9	10	3	3	3	3	1	3	2	3	3	3	0	3	83
6	8	10	10	9	10	10	3	3	3	2	0	0	3	3	3	1	2	3	83
7	9	10	8	9	10	9	3	3	3	2	2	2	3	3	3	3	3	3	88
										1			1	1	1	1		1	
Total	60	59	41	58	65	54	20	18	18	4	9	10	7	8	4	2	11	6	

[a] These categories are rated out of a possible maximum of 10 points.

[b] All these categories are rated for a maximum of 3 points.

On the basis of the total implementation score, School 7 had the highest level of implementation among the seven schools in the treatment and School 2 had the lowest. School 4 was second highest in implementation and Schools 5 and 6 were tied for the third highest degree of implementation. Schools 1 and 3, respectively, were the second and third lowest in implementation of M.A.P. Given these standings of the schools for the implementation of the program, is there any discernible relationship between implementation level and effects of program?

The most obvious pattern is that School 7 had both the highest level of implementation and the highest performance over the three grades. Can we attribute its high performance to fidelity of implementation of M.A.P.? We believe not. First, School 7 was the highest performing school from the outset, and the evidence suggests that School 7 would have performed even better had it not been doing M.A.P. Second, by Grade 2, School 7 was showing the least improvement between fall and spring of any school, and by Grade 3 was showing the greatest loss spring over fall. Third, the implementation level of School 7 was almost equal to that of Schools 4, 5, and 6, even though its performance was consistently better than those schools. Again, this result cannot support a conclusion that ties level of implementation to level of performance. Finally, the school with the worst performance overall, School 1, was not the school with the lowest level of implementation.

The teacher preparation in M.A.P. points to a similar picture: it is not possible to tie preparation to success or lack of success. In four of the seven schools, the teacher preparation was rated at the maximum by Grade 3, the grade when students overall performed at their worst. Performance in these schools was, nevertheless, widely divergent despite common levels of teacher preparation in the M.A.P. program. School 1 with the worst performance, nevertheless had teachers who were better trained in the program than School 2, which outperformed it.

LIMITATIONS

Although these and other limitations of the study have been previously discussed, it is important to remind the reader of the following when considering the conclusions and recommendations:

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 similarity of student population and school characteristics. This limitation affects the external validity of the study and thus inferences beyond the actual students studied warrant caution.

2. The attrition from about 230 children at the outset of the study to about 130 at the end could have influenced results. As discussed above, the attrition does not

affect the internal validity of the study because of the fact that the control group was selected after data collection was completed for the treatment group. It does bear upon the external validity for the same reasons discussed in limitation one.

3. Measures of program implementation were based upon self-reports and not upon direct observation by a neutral observer. Neither approach is infallible. The limitation of self-reports is that teachers may have a biased judgement of their own teaching. The strength of self-reports is that they refer to the entire period of teaching as contrasted with the limited time a neutral observer can spend in the classroom. Given the range of implementation reported, our confidence in the teachers' self-reports is increased. There appeared to be no systematic inflation in their reporting. Evidence of teachers' acceptance level of the philosophy of the M.A.P. program was not available.

6 Analysis of the M.A.P Program

On the first page of *Literacy M.A.P.*, the authors state that the 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* (National Institute of Child Health and Human Development, 2000) "provides the findings to support teaching practices outlined in this resource" (Hunter & Robinson, 2002). On the following page Hunter and Robinson report the use of a much earlier foundation for the resource, particularly the practices of "teacher directed whole group instruction, a multi-sensory approach, and the sequential development from graphemes to spelling to writing and then reading" (p. 4).

Despite stating that the NRP report supports M.A.P. teaching practices, discrepancies between the M.A.P. foundation statements and the NRP report are apparent. The NRP panel's meta-analysis concluded that phonological awareness (PA) instruction carried out in small groups was more effective than individual or whole group instruction. There was no mention of the usefulness of a multi-sensory approach to PA and phonics in the NRP report, nor does it appear among the "Phonemic Awareness Instruction" quotations from the NRP report outlined on page nine of M.A.P. Neither the cited sequential development from the NRP report nor the types of PA training cited is reflected in M.A.P. activities. There are no PA activities in M.A.P., if the common definition of phonological awareness is used; that is, instruction focused on developing an understanding of sounds in *spoken* language. Instead, early M.A.P. instruction emphasizes printing skills and learning the sounds and rules associated with each printed letter.

Also ignored in the manual is the length-of-instruction variable analyzed by the NRP. Following their meta-analysis, the NRP concluded that PA instruction lasting from 5 to 18 hours was more effective than that lasting a shorter or a longer time, a finding not

reported in the M.A.P. manual. Using the sample timetable outlined in the manual, it appears that children in kindergarten would receive approximately 87 hours of grapheme/spelling instruction, which in the M.A.P. program is incorrectly equated with PA, and additional such instruction in subsequent grades.

The NRP recommendation that teachers should help children apply PA skills in their daily reading and writing activities appears in the M.A.P. manual, as does the NRP reminder that PA does not constitute a complete reading program. However, there are no instructions in the manual for applying the letter-sound lessons being taught in daily reading as children begin this literacy program. Instead, teachers are instructed that "Graphemes begin the reading process and are taught in isolation" (p. 27; see Appendix C). Parents, too, in Correspondence Home letters, are informed that "Students are taught the graphemes (letter/sound combinations) in isolation" (pp. 90, 94, 100). Although the M.A.P. authors write, "Good quality children's literature is recommended for any reading program" (p. 28), there is no indication in the manual of how or when such reading should take place. As indicated above, the M.A.P. authors state that graphemes begin the reading process and are taught in isolation and that "[i]t is critical to students' success with reading that they have mastered the graphemes" (p. 27). A diagram of the reading process shows graphemes leading to spelling words, then dictated sentences, followed by dictated paragraphs and stories, reproducible stories, and finally children's literature. These exercises appear quite removed from the NRP recommendation that PA skills be applied in daily reading tasks.

M.A.P. activities do not appear to reflect another NRP statement listed in the manual: "PA instruction should be as relevant and exciting as possible so that the instruction engages children's interest and attention in a way that promotes optimal learning." Instead, the manual acknowledges in instructions for teachers and letters to parents that the prescribed exercises are hard work, may seem a burden, and may not be viewed as fun by children.

These comparisons make it difficult to understand how the NRP findings support the teaching practices outlined in M.A.P. as claimed in the introduction of this teaching resource. A reading program designed according to the NRP statements quoted in M.A.P., would look very different from the M.A.P. program. After quoting rather extensively from the NRP report (a report faulted by some in the research community for placing too much emphasis on synthetic PA and phonics), the authors of M.A.P. have not provided statements to justify their own approach.

Another way to examine the types of evidence the authors of the M.A.P. program might have used to create their program is to compare M.A.P. activities to findings and theory regarding PA and phonics found in the research literature. First, as stated above, there does not appear to be any PA instruction in M.A.P., because all early instruction is firmly tied to writing letters and listening to and repeating back their associated sounds.

However, as the authors of the M.A.P. manual refer to PA instruction, we will compare M.A.P. activities to research findings as if M.A.P. did contain PA instruction.

In *Preventing Reading Difficulties in Young Children,* Snow et al. (1998) wrote that a basic awareness of the phonemic structure of language is fundamental to understanding the sounds in words. A number of reading researchers (e.g., Adams, 1990; Anthony et al., 2003; Gillon, 2004; Goswami & Bryant, 1990) have recommended a PA instructional sequence that moves from developing students' awareness of speech sounds, to word-syllable-letter awareness, onset-rime awareness, and then phoneme-level skills such as blending, segmenting and manipulating. However, the M.A.P. program does not include sound awareness activities as those outlined above. Rather, sounds are tied to letters, with a focus on their correct formation.

In the literature, Blachman (2000), Bus and van IJzendoorn (1999), Byrne and Fielding-Barnsley (1989), Ehri et al. (2001) and Wagner et al. (1994) report that learning to read helps develop children's PA. Perfetti et al. (1987) refer to this as reciprocity, hypothesizing that learning to read increases reflective phonemic awareness that then promotes further gains in reading ability. This view, however, is not reflected in the M.A.P. program. Instead, reading is introduced rather late in the process; children are given books to read only after they can print and pronounce 54 graphemes, and have learned to spell dictated words, sentences, paragraphs, and stories.

Research also indicates that although PA instruction has been shown to teach PA very effectively, such instruction has only a small-to-moderate impact on reading comprehension (Bus & van IJzendoorn, 1999; Ehri et al., 2001). Both studies also note that PA is not the sole key to reading success. Thus, PA is seen to be but one component among a complex array of literacy experiences that help children learn to read. Statements in the M.A.P. manual suggest that the authors view their initial grapheme and spelling activities as a means to teach children how to be successful readers. M.A.P., however, presents a very limited array of early literacy experiences beyond those outlined above.

As M.A.P. stands for Meaningful Applied Phonics with a subtitle of Explicit Phonics through Direct Instruction, there is no doubt that the authors consider this a phonics-based program. Thus, comparisons between M.A.P. activities and phonics theory and research findings also need to be made.

First, research suggests that reading is a complex set of skills that must function together for a reader to be able to successfully comprehend text.

As reported in the literature review, Juel and Minden-Cupp (2000) found "phonics first and fast" helped low-reading group children make substantial gains in successful word reading. M.A.P. definitely endorses a "phonics first" philosophy, but their instructional

sequence does not suggest "phonics fast." Even after children start to read their own books, M.A.P. specifies that phonics instruction continue to occupy one-third of the time allotted for reading. Instruction continues to be whole group, as well. This contrasts with a widely supported recommendation that individual student differences be taken into account when planning phonics instruction, because not all students benefit equally from the same instruction (Ehri et al., 2001; Goldenberg, 2001; Pressley, 2002; Pressley, Allington, Wharton-McDonald, Block, & Morrow, 2001; Snow et al., 1998). The NRP, too, recommends flexible pacing and flexible grouping of students by phonetic ability.

M.A.P. activities do correspond to research findings on several points. First, M.A.P. instruction focuses students' attention on letter-sound correspondences in words to help them develop an awareness of the structure of words and detect patterns useful for recognizing words efficiently. As well, M.A.P. instruction is systematic, it is organized and it is specifically phonics.

However, on several points not previously outlined, M.A.P. differs from recommended practice. For example, while Stahl et al. (1998) note that good phonics instruction focuses on words, not rules, numerous rules for children to write down, memorize, and recite are listed in the M.A.P. manual. Writing using invented spelling is another widely recommended practice as it engages children in phonemic tasks (Adams, 1990; Craig, 2003; Richgels, 2001; Snow et al., 1998; Stahl, 1992). There is no invented spelling in M.A.P., where the emphasis is on learning to spell words correctly. We find no explicit recognition in M.A.P. that skillful, knowledgeable teachers who are able to identify and implement instructional approaches and practices as the situation demands are the key to reading success. Nor is there recognition that researchers have not been able to identify any one systematic phonics program that was appreciably more effective in teaching children to read than other programs offering systematic instruction. Finally, letter-sound relations and correct spelling are given precedence over real writing and reading, despite an NRP caution that phonics programs that focus 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.

Based upon this analysis, some shortcomings of the M.A.P. program components are evident. These shortcomings serve as a reminder of the importance of assessing evidence-based literacy practices in programs prior to adoption.

7 Conclusions and Recommendations

The five questions we were asked to address in this study were: (a) Which schools achieved the strongest gains over three years? (b) Are (i) higher and lower fidelity of method implementation, and (ii) higher and lower quality of method training related to achievement gains? (c) Is SES related to achievement gains? (d) Is gender related to achievement gains? and (e) Using nationally and/or provincially normed standardized scores, what was the gain made at each grade across all schools?

We note that there is no way to measure achievement gains absolutely. Achievement always has to be measured with respect to something else, either a specified criterion that has been agreed or a comparison group. For the treatment schools using the M.A.P. program, we used as a criterion to the extent possible a common set of measures from Grade 1 to 3 and studied the gains and losses students made on these measures.

It is very important to understand that because all of the treatment schools were using the M.A.P. program, there is no way to attribute any observed gains on these measures to the program without comparison to some independent and comparable group of children who were not being taught using M.A.P. To make this comparison, we introduced another set of provincially and nationally normed measures on which we obtained scores both for treatment schools and for a comparable set of control schools not using M.A.P.

CONCLUSIONS

The evidence leads to the following conclusions:

School Achievement

Due to maturation and education, each school showed gains over the three years of the study. Such gains are natural and to be expected regardless of the program of instruction. What is noteworthy is that all seven treatment schools more or less kept their same relative ranking with respect to the populations used to norm the tests. In light of this constancy, it is to be concluded that no treatment school surpassed another in terms of gains, despite the fact that some schools greatly outperformed others.

Implementation Fidelity and Achievement

We found no systematic evidence to suggest that higher and lower fidelity of program implementation was related to higher and lower achievement.

Method Training and Achievement

There was no systematic evidence to claim that higher and lower M.A.P. method training was related to higher and lower achievement.

SES and Achievement

It was impossible to separate SES from School. The most standard measure of SES in research of this type is postal code. Postal code and School were completely confounded variables, such that change of school meant both change of postal code and change of SES. The differences found between schools could just as easily be interpreted as differences in SES, but there is no possibility of separating these factors and deciding which is the one that makes the difference.

Gender and Achievement

With the exception of a few cases, which are to be expected in the analysis of so much data, gender had no effect on achievement.

Treatment Achievement Versus Control Achievement

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 in Grade 1, outperformed them in both Grades 2 and 3. Clearly, the M.A.P. program did not have as beneficial effects on students' literacy achievement as the programs used in the control schools.

RECOMMENDATIONS

The results of this study suggest some implications for policy, both in the immediate context of the program under study and for the system at large:

1. Based upon the evidence produced in this study, continued use of the M.A.P. program is indefensible. This recommendation follows 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 EPSB. What this study shows is that, however well the treatment students performed in comparison to district averages, they would have performed better with the programs used in the control schools. There is an obligation to provide students the program that the evidence suggests will maximize their performance.

2. Support for research such as this that has been provided by EPSB 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. We commend EPSB for its openness to research scrutiny witnessed in this study and hope that their openness serves as a model to other school districts in the country.

3. Steps should be in place 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. This is 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 positive, and for children's sake that possibility must be kept uppermost in our minds.

4. Potential innovations 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 a uni-dimensional focus is likely to be a less effective program for emergent readers than a multi-dimensional program.

This particular study will doubtless encourage further vigorous and important debate about reading and reading methodology – especially for diverse and challenging populations. It will also prompt further comparative research inquiry which may yield richer knowledge of program impacts on student learning. 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.

All of what we have said comes down to the support of an evidence-based practice in education. This entails resisting credulity, and accepting that even though something seems to make sense, it still might not work. A policy to adopt is to ask the question: What is the evidence?

REFERENCES

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

Anthony, J.L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38,* 470-487.

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.

Campbell, D. T. & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research.* (pp. 5-12). Chicago: Rand McNally.

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

Chew, J. (1997). Traditional phonics: What it is and what it is not. *Journal of Research in Reading, 20,* 171-183.

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

Craig, S. A. (2003). The effects of an adapted interactive writing intervention on kindergarten children's phonological awareness, spelling, and early reading development. *Reading Research Quarterly, 38,* 438-440.

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. & Cunningham, J.W. (2002). What we know about how to teach phonics. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction (*pp. 87-109). Newark, DE: International Reading Association.

Dosdall, E. (1999, December 14). *Letter to the Board of Trustees*, Edmonton Public School Board.

Duffy, G. G., & Hoffman, J. V. (1999). In pursuit of an illusion: The flawed search for a perfect method. *The Reading Teacher, 53,* 10-16.

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.

Flippo, R. F. (Ed.). (2001). *Reading researchers in search of common ground.* Newark, DE: International Reading Association.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read; Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 37-55.

Fry, E. (2004). Phonics: A large phoneme-grapheme frequency count revised. *Journal of Literacy Research, 36,* 85-98.

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

Goldenberg, C. (2001). Making schools work for low-income families in the 21st century. In S. B. Neuman & D. K. Dickenson (Eds.), *Handbook of early literacy research* (pp. 211-231). New York: Guilford Press.

Goodman, K. (1993). *Phonics phacts.* Portsmouth, NH: Heinemann.

Goswami, U. (2000). Phonological and lexical processes. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & P. Barr (Eds.), *Handbook of reading research: Vol. III* (pp. 251-267). Mahwah, NJ: Lawrence Erlbaum Associates.

Goswami, U., & Bryant, P. E. (1990). *Phonological skills and learning to read.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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 ThomsonCompany.

Hoffman, J. V. (2000). Our professional stance. In J. V. Hoffman, J. F. Baumann, & P. Afflerbach (Eds.) (with A. M. Duffy-Hester, S. J. McCarthey, & J. M. Ro), *Balancing principles for teaching elementary reading* (pp. 1-9). Mahwah, NJ: Lawrence Erlbaum Associates.

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

International Reading Association. (1999). Methods of beginning reading instruction: A position statement of the International Reading Association. [On-line]. Available:http://www.reading.org/downloads/positions/ps1033_multiple_methods.pdf.

International Reading Association. (1997). Summary statement of Role of phonics in reading instruction. A position statement of the International Reading Association. [On-line]. Available: http://www.reading.org/resources/issues/positions_phonics.html .

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. (July/August, 1997) Best practices in promoting reading comprehension in students with learning disabilities. *Remedial and Special Education, 18*(4), 197-213.

Mazzoni, S. A., & Gambrell, L. B. (2003). Principles of best practice: Finding the common ground. In L. M. Morrow, L. B. Gambrell, & M. Pressley (Eds.), *Best practices in literacy instruction* (pp. 9-22). New York: Guilford Press.

Morris, D., Bloodgood, J. W., Lomax, R. G., & Perney, J. (2003). Developmental steps in

learning to read: A longitudinal study in kindergarten and first grade. *Reading Research Quarterly, 38*, 302-328.

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.

Morrow, L. M., Gambrell, L. B., & Pressley (Eds.) (2003). *Best practices in literacy instruction*. New York: Guilford Press.

Morrow, L. M., Holt, T., Sass, A. C. (2002). Phonics Instruction. In B. J. Guzzetti (Ed.), *Literacy in America: An Encyclopedia of History, Theory, and Practice.* (pp. 428-431). Santa Barbara, CA: ABC Clio.

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: http://www.nichd.nih.gov/publications/nrp/report.html.

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.

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

Rasinski, T., & Padak, N. (2004). *Effective reading strategies: Teaching children who find reading difficult.* Upper Saddle River, NJ: Pearson.

Richgels, D. J. (2001). Invented spelling, phonemic awareness, and reading and writing instruction. In S. B. Neuman & D. K. Dickenson (Eds.), *Handbook of early literacy research* (pp. 142-155). New York: Guilford Press.

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

Scholes, R. J. (1998). The case against phonemic awareness. *Journal of Research in Reading, 21,* 177-188.

Shanahan, T. (2003). Research-based reading instruction: Myths about the National Reading Panel report. *The Reading Teacher, 56,* 646-655.

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

Spache, G. D., & Berg, P. C. (1955). *The art of efficient reading.* New York: MacMillan.

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.

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

Torgesen, J. K. (1998). Instructional interventions for children with reading disabilities. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp.197-220). Timonium, MD: York Press.

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.

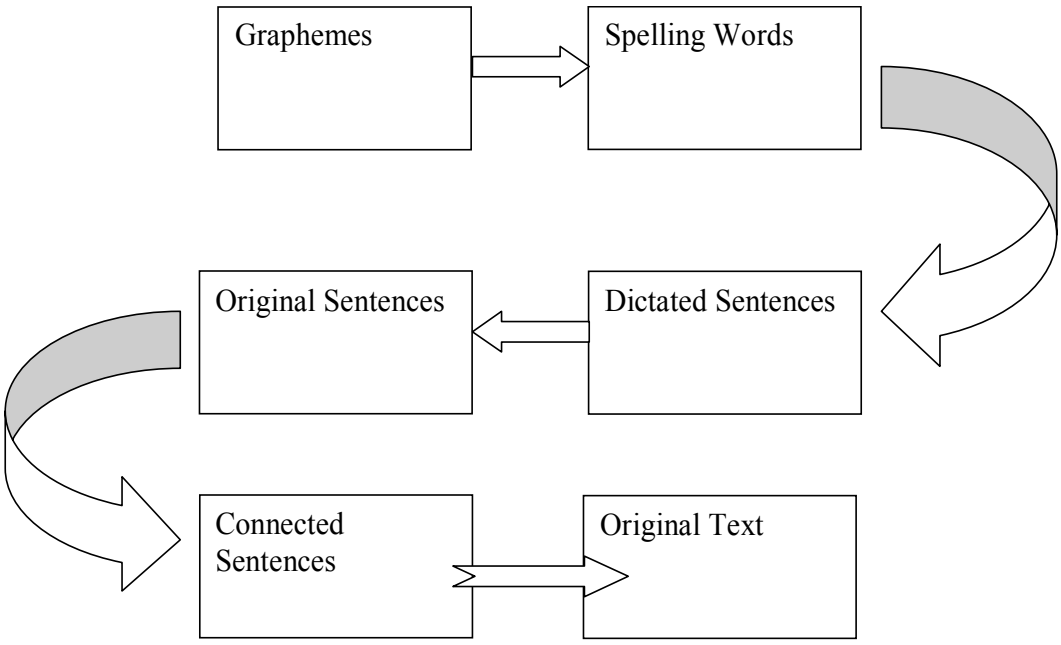
Williams, J. (1995). Phonemic awareness. In T. L. Harris, & R. E. Hodges (Eds.), *The literacy dictionary: The vocabulary of reading and writing* (p.186). Newark, DE: International Reading Association.

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

APPENDIX A

The Writing Process

Graphemes begin the writing process. This phonemic awareness activity develops the speech to print sounds of our alphabet language.



Children are guided through the writing process from teacher supported writing activities, or simulated writing experiences to writing original compositions. The goal is to ensure that children connect speech to print on four levels:

- Sounds to letters
- Letters to words (units of meaning)
- Words to sentences (words in context)
- Connecting sentences (writing on a topic)

M.A.P., p. 21

APPENDIX B

Outline of M.A.P. Lesson Models

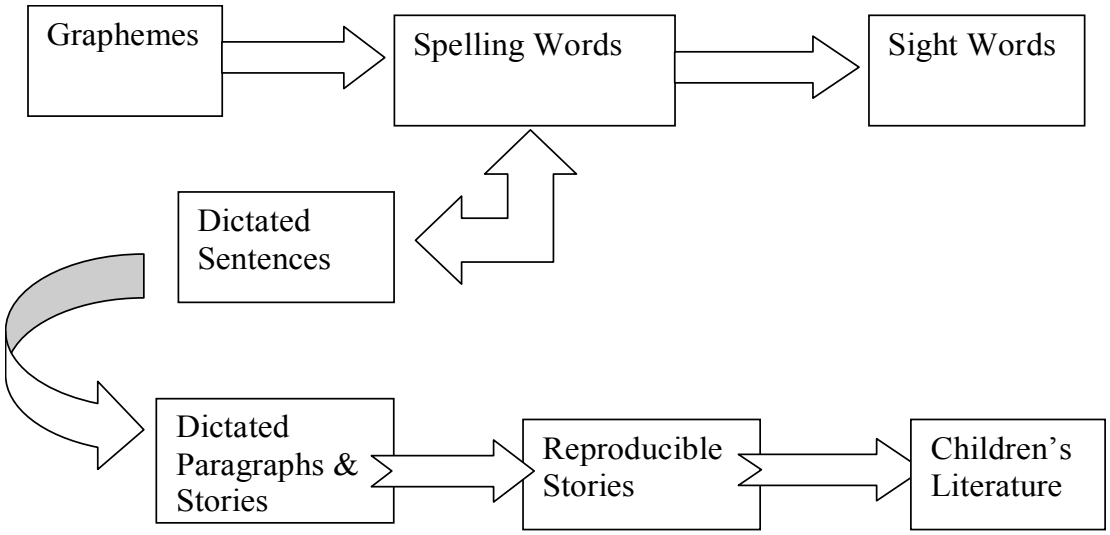
Sequence of Instruction	Explanation
Review	Begin each lesson with a quick review of previously taught relevant skills or knowledge
Goals	Ensure that students are aware of the purpose of the lesson
Scaffolding	Reduce complex tasks to smaller, more manageable steps. The students are guided through the process. This is very explicit in nature, using hints and cues to build success
Active Participation	Students are actively engaged in the learning through the multisensory approach: *see it, say it, hear it, write it* .
Check for Understanding	Question students using a variety of techniques, both whole group and individual responses.
Guided Practice	Guide students through the new learning tasks. Provide an opportunity to practice with additional support through modeling.
Systemic Feedback	During seat work, students are monitored and obtain immediate feedback.
Independent Practice	Students require time and opportunity to apply new knowledge, or to practice previously taught concepts.

Note: Explicit instruction is defined as stating directions and expectations to students in a clear and distinct manner.

APPENDIX C

M.A.P. Model of Reading Process

Graphemes begin the reading process and are taught in isolation.



M.A.P., p. 27

APPENDIX D

M.A.P. Sample Timetable

Monday	Tuesday	Wednesday	Thursday	Friday
• Review graphemes • Introduce new spelling words **(30 minutes)**	• Review graphemes • Introduce new spelling words **(30 minutes)**	• Review graphemes • Introduce new spelling words **(30 minutes)**	• Review graphemes • Introduce new spelling words **(30 minutes)**	• Quiz graphemes • Quiz spelling • Quiz grammar • Quiz reading comprehension **(45 minutes)**
• Reading • Pre-reading vocabulary development • Reading • Post-reading activity **(60 minutes)**	• Re-reading activity • Pre-reading vocabulary development • Re-reading Comprehension Instruction 1st skill practiced **(60 minutes)**	• Re-reading activity • Pre-reading vocabulary development • Post-reading comprehension instruction 2nd skill practiced **(60 minutes)**	• Writing instruction **(60 minutes)**	• Writing instruction **(45 minutes)**

Note: Intended for use in grade 1 and higher levels after spelling instruction has begun.

APPENDIX E

Thematic Vocabulary examples in Grades 1 and 2 for Science and Social Studies

Grade One	Science: Senses		
	2		
five	sens es	see	sight
hear	taste	smell	feel
2	2		2
eyes	ears	mouth	nose
fin gers			

Grade One	Science: Needs of Plants and Animals		
			3
need	plant	plants	an i mal
3 2			
an i mals	seed	liv ing	non liv ing
	3		
food	wa ter	air	shel ter

Grade Two	Social Studies: People Nearby		
3	2 2	2 2	
com mu ni ty	goods	ser vic es	food
		3	
shel ter	cloth ing	wa ter	care
	3 2		
share	to geth er	M.A.P.	leg end
north	south	east	west

APPENDIX F

Literacy M.A.P.

Teacher Survey

Use the box on the left to check the elements of the program that you use in your classroom. Additional comments can be added in the box on the right.

Phonograms	Description of activity
xxx's used here	Begin with teaching "language of instruction"(begin close to the red margin, top baseline, dotted middle line, bottom base line, in the direction we read and write…)
	Students learn the practice strokes before teaching phonograms.
	Teach phonograms 1 – 26 in isolation with printing instruction, students say the sounds orally as they write them.
	Students learn how to blend the first 26 to form one-syllable short vowel words after learning the first 26.
	Students learn phonograms 27-54 before beginning spelling. They are instructed in spacing and say the sounds while they practice writing them.
	Phonograms are reviewed daily.
	Phonograms are tested weekly.

To help determine the level of implementation, approximately how much time is alloted per day or per week for phonograms?

Per Day	Per Week

Additional comments

Use the box on the left to check the elements of the program that you use in your classroom. Additional comments can be added in the box on the right.

Spelling	Description of activity
	Say word using hands to tap out syllables
	Uses the word in a sentence
	Students repeat the word using hands to tap out syllables
	Students segment the word sound-by-sound, syllable-by-syllable orally as a group while printing the spelling word.
	Students dictate the word back and the word is written on chart paper. Word is segmented sound-by-sound, syllable-by-syllable.
	Students are asked which markings are to be used and the word is marked in red.
	Spelling rules are taught in context of the spelling word.
	Blends are highlighted and practiced
	At the end of each spelling lesson the words are read orally to develop a sight vocabulary.
	At the beginning of each spelling lesson, or throughout the day spelling words are read orally to develop sight vocabulary.
	Words are dictated in a sentence and the children print the sentence.
	Students create original sentences from spelling words taught.
	Spelling words are tested weekly.

To help determine the level of implementation, approximately how much time is alloted per day or per week for phonograms?

Per Day	Per Week

Additional comments

1. How many spelling words are you teaching/testing each week?

2. Are you using thre student reproducible charts? If yes, check off which ones and how are they used in your classroom.

Consonants/vowels		
Five Jobs for FS e		
Five sp. of er	Nouns	Verbs
Silent Final e adding suffixes	One-syllable Words with vowel suffixes	

3. Do you have a homework routine using the phonics/spelling or grammar portion of the program in place? Could you briefly describe it, and about how much time would you estimate it would take the average child in your classroom to complete?

Additional Comments:

To further assist with the collection and processing of data, please provide the following information:

1. How many hours of teacher training did you receive?

2. Did you partake in any coaching or demonstration visits? If so, how many and in what area/s?

3. How many years of experience do yuhave in using Literacy M.A.P. (or formerly, Writing Road to Reading)?

Meaningful Applied Phonics

A Longitudinal Early Literacy Study

by Linda M. Phillips, Stephen P. Norris and Dorothy J. Steffler

This Canadian study is at the centre of the "great debate" in the teaching of beginning reading. In this report, Phillips et al. of the University of Alberta Centre for Research on Literacy compare approaches to primary reading instruction in the Edmonton Public Schools District (EPSB).

Their research examines the effects of the M.A.P. program in seven schools and compares results for the M.A.P. cohort at the end of Grade 3 on a common set of literacy measures with those of a matched control group selected by the school district.

Beginning with an overview of the research literature, followed by a description of the program under study, Phillips crisply analyzes all facets of the study in accord with the data collected for the treatment group, the data provided by EPSB for the treatment and control comparisons, and the EPSB-provided data on the fidelity of program implementation. The report concludes with a set of recommendations based on the evidence.

"Despite findings which are highly controversial, the debate about reading and reading methodology among diverse student populations will be a positive outcome of this study." Professor Trevor J. Gambell, University of Saskatchewan College of Education

SOCIETY FOR THE ADVANCEMENT OF EXCELLENCE IN EDUCATION

225 - 1889 Springfield Road, Kelowna BC V1Y 5V5
Telephone 250.717.1163 Fax 250.717.1134
Email: info@saee.ca Website: http://www.saee.ca