KEVIN G. TOBIN, Pittsfield Public Schools, Pittsfield, MA

The Effects of the Horizons Reading Program and Prior Phonological Awareness Training on the Reading Skills of First Graders

Abstract: Forty students from 4 different first-grade classes in 4 different schools were identified and matched based on their Concepts About Print Test (CPT; Clay, 1970) and Phoneme Segmentation Fluency (PSF) scores from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Kaminski & Good, 1996). Two classes used *Horizons Fast Track A–B* (S. Engelmann, O. Engelmann, & Seitz-Davis, 1997) and 2 classes used a Silver, Burdett, and Ginn (SBG; Pearson et al., 1991) curriculum. Students in 1 of the *Horizons* classes and 1 of the SBG classes had received explicit phonological awareness instruction in kindergarten. Results were assessed with the Woodcock Diagnostic Reading Battery (WDRB; Woodcock, 1997), DIBELS PSF and Nonsense Word Fluency (NWF) reading tests, and measures of oral reading fluency. Results on measures of phonological skills, reading fluency, and overall reading skills indicated that classes that received instruction with *Horizons Fast Track A–B* (with and without prior phonological awareness training) significantly outperformed those students who received SBG curriculum (with and without prior phonological awareness training).

Studies sponsored by the National Institute of Child Health and Human Development (NICHD) indicated that 44% of students in fourth grade had poor reading skills (Lyon, 1998). Further, studies sponsored by the NICHD found that by the end of first grade, poor readers began to show significant decreases in self-esteem and motivation to learn to read. When these children were followed up, their problems were found to increase because they were unable to read advanced materials such as science and literature. Schools often respond to these problems by implementing expensive and time-consuming remedial reading programs. Reactive strategies that attempt to mitigate the effects of reading failure are certainly necessary; however, the only adequate solution is to prevent the development of reading problems in the first place. Research has identified at least two critical issues for preventing reading failure. The most direct strategy is to improve the quality of instruction and curriculum to teach initial reading. A second, somewhat less direct strategy is to teach critical phonological skills before reading instruction begins. This study examines the role of curriculum/instruction in reading and prior phonological awareness instruction on reading acquisition.

Phonological awareness refers to a person's awareness of the sound structure of language and involves the ability to notice and manipulate the sounds of spoken words (Mattingly, 1972). According to the National Reading Panel (2000) children who have strong

Journal of Direct Instruction, Vol. 3, No. 1, pp. 1–16. Address correspondence to Kevin Tobin at kevintobin@attbi.com.

phonological awareness tend to learn to read and spell more easily.

Chall (1967) reported the results of 25 reading investigations undertaken between 1900 and 1960. She concluded that focused instruction in phonics was superior to nonphonic instruction in teaching word recognition, oral reading, and spelling. These findings held for both low performers and typically-achieving students. Twenty-three years later, Adams (1990) reviewed the research on reading instruction and came to the same conclusion—that code-based phonics programs produced much higher reading comprehension scores than so called meaning-based programs. Similarly, Brown and Felton (1990) showed that programs that include explicit systematic phonics instruction in decoding skills, along with the opportunity to engage in meaningful reading and writing experiences, foster greater reading skill improvement than programs that do not provide explicit and systematic phonics instruction. In explicit phonics, students orally convert letters or a combination of letters into phonemes and then blend phonemes to make words. For example, when the three letters *r-a-n* are presented on the board the teacher says, "Let's see what word this is. When I point, you give me the sounds. What is it?" (Students respond "rrraaannn.") "Say it fast." (Students respond "ran.")

More recently, Foorman, Francis, Novy, and Liberman (1991) found that classes with more letter–sound instruction demonstrated improved spelling and reading skills more quickly than classes with less explicit and direct instruction. Foorman, Francis, Fletcher, Schatschneider, and Mehta (1998) found that reading programs that provide explicit and systematic instruction in phonemic decoding skills along with meaningful reading and writing experiences produced greater growth in reading skills among young children. Torgesen et al. (1999) found that explicit phonics instruction leads to greater improvement in word reading than typical basal reading programs. Thus, a long history of research including recent studies have produced a great deal of evidence that systematic explicit instruction in the phonological structure of the spoken word followed by systematic phonics offers the optimal early reading instruction for most children.

Direct Instruction is a system of teacher-directed instruction that involves teacher demonstration, careful prompting and support of student skills, and systematic correction of errors. The goal of Direct Instruction is to accelerate learning by maximizing the design and delivery of instruction (Division for Learning Disabilities and Division for Research of the Council for Exceptional Children & Tarver, 2000). All Direct Instruction programs feature scripted lessons and careful instructional sequencing. Several skills are presented in each lesson and skills are systematically reviewed. Direct Instruction differs from other explicit instruction models by emphasizing instruction in strategies and generalizations, and the importance of curriculum design.

Reading Mastery (S. Engelmann & Bruner, 1995) is the best known Direct Instruction reading series. It offers a complete elementary reading program across six grade levels. One of the key characteristics of the early reading instruction in *Reading Mastery* is the elaborate system of introducing skills and guiding students in applying these skills to a variety of words. All words encountered in connected text are decodable through the application of rules that the students have learned plus their knowledge of specific irregular words. In addition, the initial level of *Reading Mastery* includes a strong phonological component that is closely integrated with explicit instruction on phonic strategies. *Reading Mastery* is unusual in that it employs a modified orthography (letter forms) for introducing basic reading skills. These modified letter forms reduce the complication and irregularity of written English. The modified orthography is faded out after students master basic reading skills. *Reading*

Mastery is also unconventional in introducing letter sounds (that is the most common sound for each letter) before teaching letter names. This early emphasis on letter sounds (rather than names) focuses instruction on the specific skills that students must use to sound out words. Schieffer, Marchand-Martella, Martella, Simonsen, and Waldron-Soler (2002) conducted a comprehensive review of literature on the effectiveness of *Reading Mastery*. The authors examined 25 research studies that compared *Reading Mastery* or *DISTAR Reading* (a previous version of *Reading Mastery*) to other curricula. Four studies compared *Reading Mastery* to other curricula for students in general education. In two studies, *Reading Mastery* was found to be superior to other programs. Mixed results occurred in two studies involving reading readiness among general education students. Eight studies examined the effectiveness of *Reading Mastery* or *DISTAR Reading* for remedial reading students. In six of these eight studies, *Reading Mastery and DISTAR Reading* were more effective than other curricula. For the remaining two studies, no significant difference was found.

Horizons is a more recently developed Direct Instruction reading program. Like *Reading Mastery*, *Horizons* incorporates research findings concerning optimal decoding and comprehension strategies. For example, the *Horizons Fast Track A–B* program presents a diversity of story reading formats to broaden students' understanding of character, plot, and sequence. The *Horizons* reading program has three levels, A, B, and C–D. In addition, an accelerated version of the first two levels is available as *Horizons Fast Track A–B*. Unlike *Reading Mastery*, the *Horizons* reading program uses typical orthography and teaches letter names initially. These differences give *Horizons* a much more conventional appearance and sequence than *Reading Mastery*. *Horizons Fast Track A–B* was field-tested between 1992 and 1998 (SRA/McGraw-Hill, 1999) and revised four times based on detailed instructor notes and daily performance assessment of students. Assessment included Oral Reading Fluency measures. The field tests occurred in four classrooms in four different states. Students came from diverse backgrounds and lived in urban and suburban districts.

The only previous research on the effectiveness of *Horizons* was described in an unpublished study by Vreeland, Huth, Lum, Pattison, and Vail (1998). This study found that 17 students from low socioeconomic backgrounds from one class who received instruction in *Horizons Fast Track A–B* improved grade equivalent scores from 1.0 to 2.7 in Passage Comprehension on the Woodcock Reading Mastery Test—Revised (WRMT—R; Woodcock, 1998). A control group of eight students with similar backgrounds improved grade equivalent scores from K.9 to 2.2 in Passage Comprehension. However, the groups were not matched on initial reading skill so we cannot rule out initial differences in reading skill as a contributor to this difference between groups.

The purpose of the present study was to evaluate the effect of the *Horizons Fast Track A–B* program on students' decoding skills and oral reading fluency. I hypothesized that students given explicit decoding training in *Horizons Fast Track A–B* would have better word attack and word recognition skills than students in a conventional reading program by the end of first grade. Furthermore, I hypothesized that students who received phonological awareness instruction in kindergarten would have better word recognition skills than those who did not have this background.

Method

Participants

The present study was designed in collaboration with the school district's reading curriculum coordinator. The school district had decided to pilot the *Horizons Fast Track A–B* program in two first-grade classes within two dif-

ferent schools and to contrast their performance with two classes in which reading was taught using the district's current reading curriculum, SBG. The evaluation included 39 students—20 girls and 18 boys, 5 African Americans and 33 Caucasians. The proportion of African-American students (13%) to Caucasian students (86%) was approximately the same as in the city where this study took place (88% Caucasian, 9% African American). Two of the classes came from schools in more affluent sections of the city and two classes came from schools with high rates of poverty (44% and 67% free and reduced-cost lunch). Table 1 shows demographics for each school in this study.

Inclusion and Matching Criteria

In order to control differences in prior reading skill and ensure a fair comparison between instructional conditions, students were selected for the study based on their match with students in other classes. A match was defined as having a score on the CPT in the same quartile as that of a student from a different class. The CPT was administered at the end of kindergarten by reading teachers. Students were also matched based on the PSF measure from DIBELS administered in September of the students' first-grade year. All students had CPT scores. Class lists were searched to find students in each class with CPT scores that were within a few points of students in each of the other classes. Then the PSF test was given to the students who had matches on the CPT. Students were selected for the study if they matched with students in the other classes on both PSF and CPT scores. Each student selected had a comparable student in each of the other three classes with a PSF score within 8 points. Intelligence test scores, socioeconomic status (SES), gender, and race were not used to match students.

Nineteen students received instruction in *Horizons Fast Track A–B* and 19 students received instruction using SBG. Students were not randomly assigned; rather they were administratively assigned to schools, then classes within schools. They were selected for the study based on their match with students in the other classes. All students were at least 6 years of age in September of first grade. Initially, there were 20 girls and 20 boys in the study. Three students (two girls, and one boy) moved in the spring. Three replacement students with matched CPT and PSF scores were added, but one of these students also moved in the last quarter of the year and could not be replaced. One girl was absent so frequently that her scores could not be used.

Table 1

Percent of Free and Reduced-Cost Lunch and the Percent of Students in Ethnic Groups by Schools

School	Reading Program	PA Training	FRL	%Black	%Hispanic	%White	%Asian
School A	*Horizons*	Yes	44.4	12.5	3.2	80.3	3.2
School B	*Horizons*	No	22.4	5.2	0.8	90.9	2.7
School C	SBG	Yes	67.0	22.1	13.3	62.7	1.0
School D	SBG	No	17.8	2.5	0.2	96.4	0.0

Note. PA = phonological awareness.

Each reading program was implemented with one school that served a higher socioeconomic neighborhood and one that served a lower socioeconomic neighborhood. The highest SES school (School D) received the SBG reading program as did the lowest SES school (School C). Thus SES was counterbalanced with a reading program. The phonological awareness training was implemented in the two lower SES schools. Thus, each reading program (*Horizons* and SBG) was implemented with one class that had received phonological awareness training and one that had not. Since the phonological training was implemented only at low SES schools, we cannot separate the presumably negative effects of the lower SES and the effects of the phonological awareness training program.

Dependent Variable and Measures

Pretests, administered in early September, included the Concepts About Print Test (CPT), and the Phonemic Segmentation Fluency (PSF) and Nonsense Word Fluency (NWF) tests from the DIBELS. A midtest of Oral Reading Fluency (ORF) was given in January. Posttests on PSF, NWF, ORF, and the Woodock Diagnostic Reading Test were administered in May. Thus, outcome variables included phonological awareness, decoding skill, oral reading fluency, and broad reading skills.

Concepts About Print Test. The CPT is an informal reading assessment developed by Clay (1970). CPT uses trade books and has kindergarten students demonstrate their awareness of reading and letters while sitting with a reading teacher. It consists of asking children if they know where to begin reading a book and if they know that reading progresses from left to right. It has the face validity of having a kindergarten child demonstrate knowledge of reading behavior. As was noted above, the CPT was given in spring of the students' kindergarten year and was used only for initial matching of students across classrooms.

DIBELS. The DIBELS consists of brief measures of critical early literacy skills, including Phoneme Segmentation Fluency, Nonsense Word Fluency, Letter Naming Fluency, and Onset Sound Fluency. Good, Simmons, and Smith (1998) found that the DIBELS is an effective measure for evaluating the acquisition of early reading skills.

The two DIBELS measures used in this study were PSF and NWF. In the PSF measure, the experimenter says a list of up to 24 words, one at a time, to a child and asks the child to say all the sounds that they hear in each word. For example, the examiner says "frog" and the child says /f/ /r/ /o/ /g/. As soon as the student completes the response, the examiner says a new word. This continues for a period of 1 min. The score is the total number of phoneme segments identified in 1 min. Students were tested individually. The author administered the DIBELS and ORF assessments in a quiet hallway or office in the school. Twenty alternate forms of the PSF test were available. Alternate form reliability for individual probes was estimated at .88 and for three probes, it was .96. Kaminski and Good (1996) found that PSF scores correlated with scores from the Woodcock-Johnson Psycho-Educational Battery (Woodcock & Johnson, 1989) readiness cluster score at .54.

The NWF test of the DIBELS presents students with a sheet of 80 vowel–consonant or consonant–vowel–consonant nonsense words (i.e., ef, lut). Students were trained on practice items and then were told, "When I say 'begin' read the words as best you can. Point to each letter and tell me the sound or tell me the whole word." The score was the number of letter sounds correctly identified in 1 min. NWF has a 1-month alternate form reliability of .83. The concurrent criterion validity of NWF with the Woodcock-Johnson Psycho-Educational Battery readiness cluster was .59. Neither *Horizons Fast Track A–B*, nor SBG taught nonsense words. Therefore, NWF constitutes a measure of generalization

of decoding skill that is not closely related to either curriculum.

Oral Reading Fluency. ORF is a 1-min oral reading test with several alternate forms. The first-grade passages were derived from a 1984 version of the Houghton-Mifflin basal reader. Each administration involved a different passage. Hintze, Owen, Shapiro, and Daly (1998) selected these probes by using a random number generator to indicate the page from which the passage should be selected (see Shinn, 1989, for standard procedures). Two scores were derived from these probes, correctly read words per minute (CWPM) and error words per minute (EWPM). All probes for the first-grade level had a readability score between 1.0 and 1.9. These probes were used because they were grade leveled and were neutral with respect to the curricula taught in this study. Students were given one passage to read while the examiner had a second copy of the passage with a cumulative word count in the right hand margin. Hintze et al. reported parallel forms reliability for these probes ranging from .72 to .96. Test–retest reliability ranged between .92 to .97. The alternate forms reliability of this measure ranged from .89 to .94. Criterion related validity ranged from .52 to .91 with the Stanford Diagnostic Reading Test (Harcourt Brace and Co., 1996), the PIAT (Dunn & Markwardt, 1989), and the WRMT—R.

Woodcock Diagnostic Reading Battery. The Basic Reading skill index score from the WDRB was used at the end of the school year to assess final reading proficiency. Internal consistency reliability ranged between .92 to .95. Test–retest reliability ranged between .90 and .94. The correlation of the Basic Reading skill index of this measure was .80 with the PIAT, .64 with the Reading Comprehension subtest of the K-TEA (A. S. Kaufman & N. L. Kaufman, 1985), and .86 with the WRAT—R (Wilkinson, 1993). The Basic Reading score is a composite that summarizes the Word Attack, Letter–Word Identification, and Passage Comprehension subtests. An experienced reading teacher was hired as an independent evaluator. She assessed the children with the Woodcock at the end of the year and scored the protocols independently with the appropriate scoring software. She was blind to the instructional conditions in this study and to performance on other measures used in this study.

Materials

Phonological awareness training. Telian's *Lively Letters* (Telian, 1997) is based on the work of Lindamood and Bell. It teaches letter sounds by having cartoon-like letters say their names. Students are taught to identify the position of the tongue and the shape of the mouth needed to make each sound. The program attempts to teach students to recognize the connection between sound, mouth movements, and letter shapes in a structured way. The *Lively Letters* has mnemonic picture stories and hand cues to facilitate rapid retrieval of letter sounds and improve decoding. Teachers taught 44 letter–sound associations explicitly through intense use of mnemonics and imagery. Initial sounds instruction was followed by work on single syllables. There was an emphasis on self-cueing and rapid naming. This curriculum was used to teach phonemic awareness in kindergarten for one class that received *Horizons Fast Track A–B* and one class that received SBG. Two of the classes, Class A and Class C, received at least 40 min per day of phonemic awareness training with the Telian's *Lively Letters* to enhance phonological awareness for 6 months in kindergarten. One kindergarten teacher and one special education teacher taught *Lively Letters*. They taught in different schools. No other explicit phonemic awareness training was done. The students who entered the other two classes, Class B and Class D, did not have explicit training in phonemic awareness in kindergarten.

Horizons Fast Track A–B. The *Horizons Fast Track A–B* program provides structured teaching of reading skills. It has a track design, which presents four or five skills per lesson. Across

lessons, skills become more sophisticated and students apply skills increasingly independently. *Horizons Fast Track A–B* teaches phonics explicitly and systematically. It has procedures for dealing with phonologically irregular words. There is considerable opportunity to read decodable text. Spelling exercises reinforce the relationship between sounds and spelling patterns. Activities are used to help students increase reading comprehension. Each lesson has five parts: (a) letters and sounds instruction and practice, (b) word attack skills instruction and practice, (c) oral reading of a story, (d) story-based activities such as independent workbook activities, and (e) letter writing, sentence writing, and spelling. *Horizons Fast Track A–B* has scripted lessons in a teacher presentation book. The first 55 lessons are highly prompted.

Teachers who used the *Horizons Fast Track A–B* program were selected on the basis of their willingness to pilot the material for 1 year. The teachers were trained by a trainer from SRA/McGraw-Hill on how to use the program. Teachers were given 1 day of training with follow-up consultation (verbal and written) every 3 months by the initial trainer. The training consisted of oral explanations of the curriculum, consultant modeling, and guided teacher practice. Teachers were trained how to follow the script, how to do correction procedures, and how to maintain the pace of instruction. The two teachers completed all 150 lessons. They went on to do 15 lessons in *Horizons Fast Track C–D* at the end of the year.

Silver, Burdett, and Ginn. The SBG program has been the district's reading curriculum for 15 years. It consists of an anthology of children's literature by popular children's authors. In subsequent revisions, phonics activities were added, but these activities were not coordinated with the words that appear in the program's stories. Each lesson includes silent reading and workbook activities. Phonic skills are taught in three stages. In the first stage, students are taught letter shapes, names, and sounds. Teachers encourage students to think of a word that begins or ends with a particular sound. At the second stage, more letter patterns are introduced, including consonant clusters, vowel digraphs, and phonograms. Children are encouraged to guess at unknown words. At the third stage, multiple-syllable words are introduced. The district reading coordinator selected the control classes. The reading coordinator was responsible for monitoring the SBG instruction to assure its implementation. All teachers followed the scope and sequence specified by the publisher. These teachers were initially trained to teach using the SBG program several years previously by the publisher. Instruction in SBG is not scripted, and considerable teacher discretion is involved in deciding the pace and order of the lessons.

Results

The dependent measures were analyzed with a series of 2×2 (curriculum $\times$ prior phonological training) ANOVAs. In each case, the main effect for curriculum (*Horizons* vs. SBG), the main effect for prior phonological training (training vs. no training), and the interaction of these two variables was examined. In addition to the traditional test of statistical significance, measures of effect size were computed. All effects were described by eta^2, a measure of the percentage of variance in the outcome that is statistically associated with the effect being tested. The magnitude of main effects was also described by standardized mean differences (*SMD)*. The *SMD* for the main effect of curriculum was computed as the difference between the mean of the combined *Horizons* group (i.e., combining those with and those without phonological training) and the combined SBG group, divided by the combined standard deviation. The *SMD* for the main effect of phonological awareness training was computed similarly. This pattern of analysis was repeated for (a) the two pretests—DIBELS, PSF, and NWF; (b) the two midtests—oral reading rate of correct words

and errors; and (c) the five posttests—DIBELS, PSF, and NWF, oral reading rate of words correct and errors, and the WDRT Basic Reading Skills scale. MANOVA was not used because of small sample size. Beyond these statistical analyses, plots of individual scores of ORF and WDRT were analyzed.

Pretests of DIBELS (PSF and NWF)

Pretreatment differences among groups, shown in Table 2, were assessed with two analyses of variance. On the PSF test, the main effect for curriculum (*Horizons*/SBG) was not statistically significant, $F(1, 34) = 2.84$, $p = .10$, $eta^2 = .08$, $SMD = 0.6$; nor was the main effect for phonological awareness training, $F(1, 34) = 1.39$, $p = .25$, $eta^2 = .04$, $SMD = 0.1$; nor was the interaction between curriculum and phonological training, $F(1, 34) = 1.39$, $p = .33$, $eta^2 = .03$. Note that although the main effect was not close to the level of statistical significance, the difference between the curriculum groups was noticeable with an *SMD* of 0.6. On the NWF test, the main effect for condition was not statistically significant, $F(1, 34) = 2.06$, $p = .16$, $eta^2 = .06$, $SMD = 0.6$; nor was the main effect for phonological training, $F(1, 34) = 1.95$, $p = .17$, $eta^2 = .05$, $SMD = 0.7$; nor the interaction, $F(1, 34) = 0.36$, $p = .55$, $eta^2 = .01$. Similar to the previous analysis, although not close to achieving statistical significance, there were noticeable differences among groups on the pretest of NWF. In this case, differences are evident in both main effects. Across the six tested effects related to pretest differences among groups, no p value was less than .10. Thus, differences are similar to those that would be expected as a result of random variability.

Midtests of Oral Reading Fluency

Results of the tests of ORF, given halfway through the school year (early January) are shown in Table 3 and were analyzed with two 2×2 ANOVAs. The results indicate that there was a statistically significant effect of curriculum, $F(1, 34) = 6.00$, $p = .02$, $eta^2 = .15$, $SMD = 0.7$; on CWPM. There was also a

Table 2

Pretest Results

	No Phon. Training		Phon. Training		Overall	
	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)
	Phoneme Segmentation Fluency					
Horizons	19.3	(12.4)	28.0	(13.0)	23.9	(13.1)
SBG	16.6	(10.8)	17.3	(12.8)	16.9	(11.5)
Overall	17.9	(11.5)	22.7	(13.7)	20.4	(12.7)
	Nonsense Word Fluency					
Horizons	24.1	(12.5)	26.1	(3.4)	25.1	(9.0)
SBG	19.0	(4.0)	24.0	(6.9)	21.5	(6.1)
Overall	21.7	(9.6)	25.1	(5.3)	23.4	(7.9)

statistically significant main effect for prior phonological awareness training, $F(1, 34) = 5.23$, $p = .03$, $eta^2 = .13$, $SMD = 0.7$. The interaction approached, but did not achieve statistical significance, $F(1, 34) = 3.74$, $p = .06$, $eta^2 = .10$. Thus, by January there were large and statistically significant differences among the classes in number of CWPM. In particular, the two *Horizons* classes scored substantially above the two SBG classes, and classes that had experienced prior phonological training scored above those that had not received this training. The sizable, though not statistically significant interaction indicates a trend in the direction of prior phonological training being particularly important when it is combined with the *Horizons* curriculum. This interaction is seen most dramatically in the comparison between the reading rate of the group that received *Horizons* after phonological awareness training (mean of 66.9 CWPM) and the group that received SBG without previous phonological training (mean of 20.9 CWPM).

Students' error rates were also analyzed. The main effect for curriculum was statistically significant, $F(1, 34) = 51.26$, $p < .001$, $eta^2 = .60$, $SMD = 1.51$. The main effect for phonological training was not statistically significant, $F(1, 34) = 2.43$, $p = .13$, $eta^2 = .07$, $SMD = 0.33$; nor was the interaction, $F(1, 34) = 0.45$, $p = .51$, $eta^2 = .01$. The main effect of curriculum is strikingly large accounting for fully 60% of the variance in students' reading errors. This indicates that students in the *Horizons* groups made substantially fewer errors (mean of 2.8 errors per minute) than those in the SBG group (mean of 7.9 errors per minute).

Posttests of DIBELS (PSF and NWF)

Posttest results are presented in Table 4. The results of the DIBELS posttests were analyzed in two 2×2 ANOVAs, one for scores on PSF and one for scores on NWF. The results of the PSF tests show a very large and statistically significant main effect for curriculum, $F(1, 34) = 43.54$, $p < .001$, $eta^2 = .56$, $SMD = 1.5$; and small, statistically nonsignificant

Table 3

Midtest Results

	No Phon. Training		Phon. Training		Overall	
	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)
	Passage Reading CWPM					
Horizons	25.9	(25.6)	66.9	(46.0)	47.5	(42.3)
SBG	20.9	(14.3)	24.3	(23.0)	22.5	(18.5)
Overall	23.3	(20.0)	46.7	(42.1)	35.0	(34.6)
	Passage Reading EWPM					
Horizons	3.7	(2.5)	2.1	(1.7)	2.8	(2.2)
SBG	8.2	(2.3)	7.6	(2.2)	7.9	(2.2)
Overall	5.9	(3.3)	4.9	(3.4)	5.4	(3.3)

differences across levels of phonological training, $F(1, 34) = 0.46$, $p = .83$, $eta^2 < .001$, $SMD = 0.04$; and interaction, $F(1, 34) = 1.55$, $p = .22$, $eta^2 = .04$. Students in the *Horizons* curriculum scored well above those in SBG. The difference between the programs was associated with 56% of the variance in scores, and the *SMD* was a very large 1.5. In contrast, by the end of the first-grade year, the prior phonological training accounted for little variance (1%) and this prior training did not appear to interact with the reading instruction condition (interaction accounted for 4% of the variance in scores). On average, students in the *Horizons* program were able to identify approximately 65 phonemic segments per minute and the students in SBG were able to identify only 46.

Similarly, analysis of NWF posttests revealed a large and statistically significant main effect of curriculum, $F(1, 34) = 10.36$, $p = .003$, $eta^2 = .23$, $SMD = 0.94$; with small and statistically insignificant effects for phonological training, $F(1, 34) = 0.96$, $p = .33$, $eta^2 = .03$, $SMD = 0.30$; and interaction, $F(1, 34) = 0.43$, $p = .52$, $eta^2 = .01$. The pattern closely mirrors that seen with the PSF measure. The magnitude of effects associated with the *Horizons* curriculum is very large, accounting for 23% of the variance in test scores. The *Horizons* group read an average of 102 nonsense words per minute and the SBG group read only 55 per minute.

Posttests of Oral Reading Fluency

Posttests of ORF are shown in Table 4. These results were analyzed in two 2×2 ANOVAs—one of these analyzed CWPM and the other analyzed EWPM. The results indicate that on CWPM, there was a large and statistically significant effect of curriculum, $F(1, 34) = 27.31$, $p < .001$, $eta^2 = .45$, $SMD = 1.29$; and small, statistically nonsignificant main effects of phonological training, $F(1, 34) = 0.63$, $p = .43$, $eta^2 = .02$, $SMD = 0.26$; and interaction, $F(1, 34) = 1.96$, $p = .17$, $eta^2 = .06$. Thus, educational program was associated with large educationally significant differences in ORF by the end of the first grade. Students in *Horizons* (with and without prior phonological awareness) read at an average rate of 104 CWPM while the students in SBG read at only 46 CWPM. Figure 1 shows each student's performance on the oral reading test. Each dot on the figure represents the performance of an individual student and the short horizontal lines represent group means. This figure makes the large magnitude of mean differences among groups clear. Also striking is the fact that not one student in either *Horizons* group read below 60 CWPM and only the top four students in the SBG groups achieved this rate.

The analysis of oral reading errors per minute found a large and statistically significant main effect for curriculum, $F(1, 34) = 21.56$, $p < .001$, $eta^2 = .39$, $SMD = 1.22$; the main effect of phonological training was not statistically significant, $F(1, 34) = 0.52$, $p = .70$, $eta^2 = .01$, $SMD = 0.1$; and the interaction was not statistically significant, $F(1, 34) = 1.18$, $p = .29$, $eta^2 = .03$. Thus, the educational program had substantial educational significance in terms of decreased reading error rate in first grade. The *Horizons* classes with and without prior phonological awareness averaged approximately 0.9 errors per minute while the SBG classes averaged 3.7; the difference between groups was associated with fully 39% of the variance in scores.

Effects of Instruction on Woodcock Diagnostic Reading Test

The outcomes for each group from the WDRT—R are shown in Table 4. A 2×2 ANOVA was used to evaluate the effects of program and prior phonological awareness training on WDRT Basic Reading scores. A large and statistically significant effect was evident in the main effect of curriculum, $F(1, 34) = 7.32$, $p = .011$, $eta^2 = .18$, $SMD = 0.86$. The groups that experienced the *Horizons* program received an average score of 121 on this individually administered standardized test while those who

received SBG averaged 112. In contrast, prior phonological awareness training did not produce a significant effect on WDRT Basic Reading scores, $F(1, 34) = 0.82$, $p = .37$, $eta^2 = .02$, $SMD = 0.36$; nor did the interaction between curriculum and phonological training, $F(1, 34) = 0.36$, $p = .54$, $eta^2 = .01$. These results are shown with greater detail in Figure 2. This fig-

Table 4

Posttest Results

	No Phon. Training		Phon. Training		Overall	
	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)
	Phoneme Segmentation Fluency					
Horizons	64.0	(9.7)	67.1	(7.8)	65.6	(8.6)
SBG	47.9	(7.8)	43.5	(11.2)	45.6	(9.7)
Overall	55.9	(11.9)	55.3	(15.3)	55.6	(13.6)
	Nonsense Word Fluency					
Horizons	90.7	(38.0)	115.0	(53.2)	102.9	(46.6)
SBG	52.6	(35.0)	57.4	(53.6)	55.0	(44.0)
Overall	72.6	(40.6)	87.7	(59.7)	80.2	(50.9)
	Passage Reading CWPM					
Horizons	91.1	(30.0)	115.4	(46.8)	103.9	(40.6)
SBG	48.8	(28.3)	42.1	(25.7)	45.6	(26.6)
Overall	68.8	(35.7)	80.7	(52.9)	74.8	(44.9)
	Passage Reading EWPM					
Horizons	1.1	(1.1)	0.7	(0.9)	0.9	(1.0)
SBG	3.2	(2.5)	4.1	(2.2)	3.7	(2.4)
Overall	2.2	(2.2)	2.4	(2.4)	2.3	(2.3)
	WDRT Standard Score					
Horizons	118.4	(8.1)	123.0	(8.8)	121.1	(8.6)
SBG	112.0	(10.8)	112.9	(9.2)	112.4	(9.8)
Overall	114.8	(9.9)	118.5	(10.2)	116.7	(10.1)

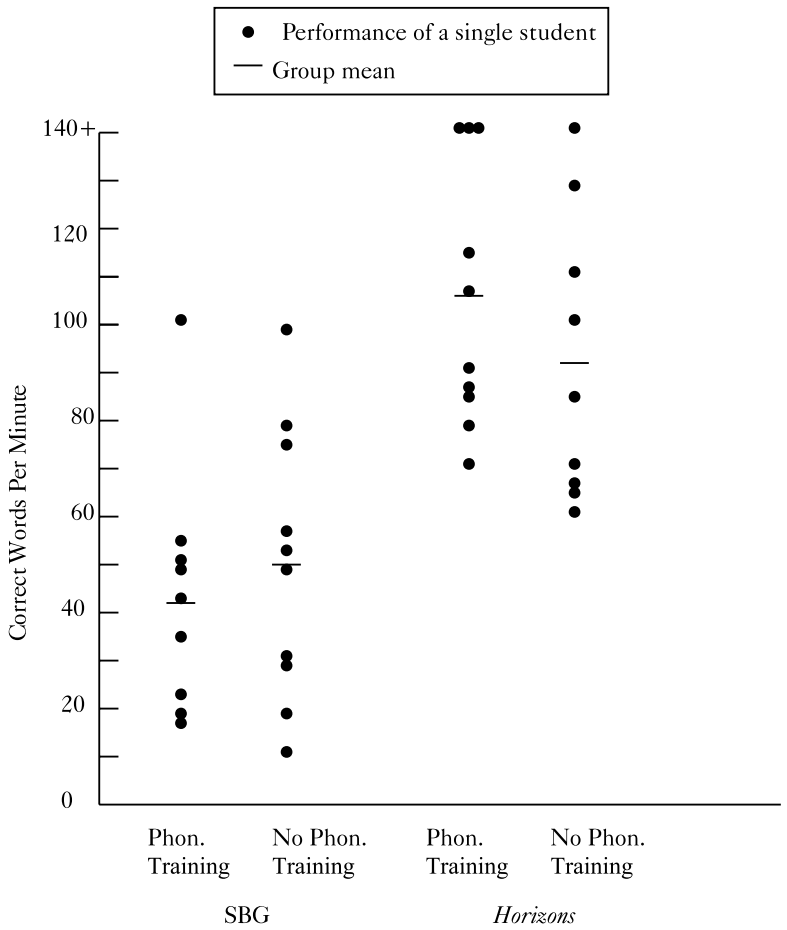
ure shows the consistency of scores in each group. Note that only 3 of the 19 SBG students scored as high as the average of the *Horizons* group that did not have phonological training and only one SBG student achieved the level of the average student in the *Horizons* plus phonological training group. Conversely, only 3 of the 19 *Horizons* students scored below the mean of the SBG groups. This is a strikingly clear pattern of differentiation between the conditions.

Discussion

Although the students in this study came into first grade with roughly comparable levels of early literacy skills, the students who received *Horizons Fast Track A–B* showed substantially stronger reading skills by the end of the year. This difference was seen across all measures used in this study including tests of PSF, NWF, ORF and accuracy, and an individually

Figure 1

Dotplot of posttest oral reading fluency results.

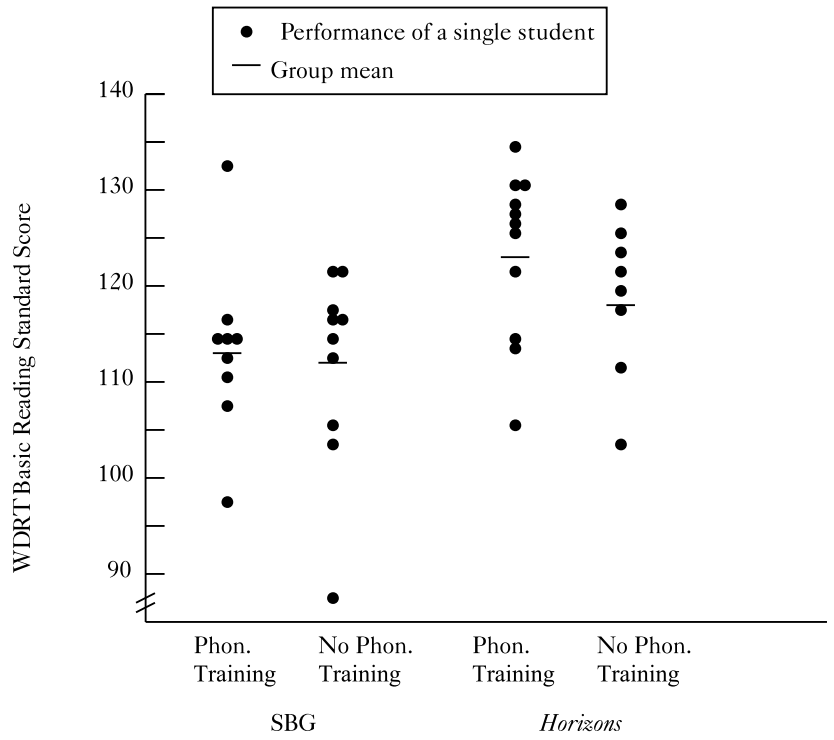


administered standardized test of basic reading skills. Students who had previous phonological training in combination with the *Horizons* program showed initially higher skill levels, but over time these students converged with those who had the *Horizons* program without prior phonological training.

These results, like those from any applied study, must be interpreted with caution. Because of the nature of public schools, random selection of teachers and students to participate in the study, and random assignment of these participants to the four treatment groups was not possible. As a result, the particular teachers and students in this study may not be representative of those in other schools. In addition, the teachers or students could have systematically differed across groups in unknown ways. Teachers who volunteered to teach *Horizons Fast Track A–B* may have been more adventuresome and more engaged than teachers who did not volunteer. Larger sample size, random assignment of subjects to conditions, and use of more classes implementing the curricula are needed before broad generalization can be made about the relative efficacy of the *Horizons Fast Track A–B* curriculum compared to SBG. In the absence of random assignment, students were assigned to schools

Figure 2

Dotplot of posttest results on Woodcock Diagnostic Reading Test.



and classrooms based on administrative issues including the location of their home. Students from these intact classrooms were selected in order to create the best possible matches across groups. The pretests, sensitive measures of important early literacy skills, found differences that were not statistically significant. However, matching was not perfect. Effect size measures showed that the two *Horizons* groups performed noticeably better on the pretests than did the two SBG groups. However, the magnitude of differences between the conditions on the mid- and posttests was much larger than the differences on the pretests. On all measures, there can be no question that students in the *Horizons* group opened a much larger gap over the SBG students during the course of the school year.

Although students in one SBG class had several hours of phonological awareness training in kindergarten, this was not sufficient to make their reading acquisition comparable to students instructed with *Horizons Fast Track A–B.* This kindergarten phonological awareness training did not have a positive effect on reading by the end of first grade. It appeared to result in better oral reading fluency in January of first grade, but this effect faded by May. It did not appear to have much effect on reading error rate, Woodcock Basic Reading, or nonsense word reading at the end of first grade. This lack of effect may be due to the small sample size because small sample size results in limited power to detect true differences. The lack of effect could also be a result of the Telian curriculum. It would be advisable to more thoroughly monitor the quantity and quality of phonological awareness training instruction. This assessment would allow one to specify the intensity of this type of instruction and relate it to student outcome. The Telian *Lively Letters* program is only one of many phonological awareness programs. Results of this study should not be generalized to other phonological curricula; however, these results should be taken as a reminder that implementation of a phonological awareness program should not be *assumed* to substantially improve later reading development. In addition, the findings regarding the phonological awareness program should be viewed with extra caution because this program was implemented only in the lower SES schools. The apparent effects of the program could have been reduced because the students who did not have the program would have had other advantages associated with their higher SES.

The findings of this study are somewhat inconsistent with prior research. For example, Byrne and Fielding-Barnsley (1991) found that phonemic awareness training improved children's skill to decode unfamiliar words. Foorman, Francis, Novy, et al. (1991) found that letter–sound instruction mediates progress in first grade reading and spelling acquisition. The key question is, once phonological awareness training is initiated in kindergarten, how much systematic phonics instruction is needed to maintain these skills as reading skills are developing in first grade?

It may be, as suggested by Adams (1990), that generalization of phonics skills is more difficult to obtain in a literature-based basal curriculum where vocabulary is not controlled and phonics lessons are not linked to passages of connected text. The other issue is the quality of phonemic awareness training. Although the classes were matched for the type of training provided, it was not possible to specify the exact instructional procedures or lesson length. Qualitative and quantitative factors in the provision of phonemic awareness training could have contributed to these results.

What was most striking about students in the *Horizons Fast Track A–B* program was the accuracy of their reading. Even the slowest reader in *Horizons* was very accurate and made fewer than two errors while reading between 60 and 90 words a minute. These results suggest that not all reading curricula maximize the critical components of reading, and that teachers and school administrators should examine how the

curriculum they select performs on critical benchmarks of reading proficiency. It is imprudent to select a curriculum that purports to be effective before it has been pilot tested and compared to the curriculum that is currently in use in a district. These results indicate that the *Horizons Fast Track A–B* program produces superior reading outcomes in first grade.

References

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

Brown, I. S., & Felton, R. H. (1990). Effects of instructions on beginning reading skills in children at risk for reading disability. *Reading and Writing: An Interdisciplinary Journal, 2,* 223–241.

Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 83*(4), 451–455.

Chall, J. S. (1967). *Learning to read: The great debate.* New York: McGraw-Hill.

Clay, M. M. (1970). An increasing effect of disorientation on the discrimination of print: A developmental study. *Journal of Experimental Child Psychology, 9,* 153–162.

Division for Learning Disabilities and Division for Research of the Council for Exceptional Children, & Tarver, S. G. (2000). Research alert: Focus on Direct Instruction. *Effective School Practices, 18*(3), 30–32.

Dunn, L. M., & Markwardt, F. C. (1989). *Peabody Individual Achievement Test.* Circle Pines, MN: American Guidance Service.

Engelmann, S., & Bruner, E. C. (1995). *Reading Mastery I* (Rainbow ed.). Columbus, OH: SRA/McGraw-Hill.

Engelmann, S., Engelmann, O., & Seitz-Davis, K. L. (1997*). Horizons Learning to Read Fast Track A–B*. Columbus, OH: SRA/McGraw-Hill.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, S., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 1–15.

Foorman, B. R., Francis, D. J., Novy, D. M., & Liberman, D. (1991). How letter sound instruction mediates progress in first grade reading and spelling. *Journal of Educational Psychology, 83,* 456–469.

Good, R. H., Simmons, D. C., & Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review, 27,* 45–56.

Harcourt Brace and Co. (1996). *Stanford Achievement Test Series, Ninth Edition, Technical Data Report.* San Antonio, TX: Author.

Hintze, J. M., Owen, S. V., Shapiro, E. S., & Daly, E. J. (1998). Generalizability of oral reading fluency measures: Application of G Theory to curriculum based measurement. *School Psychology Quarterly, 15,* 52–68.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy. *School Psychology Review, 25*(2), 215–227.

Kaufman, A. S., & Kaufman, N. L. (1985). *Kaufman Test of Educational Achievement.* Circle Pines, MN: American Guidance Service.

Lyon, G. R. (1998). *Overview of reading and literacy initiatives.* Retrieved May 4, 2000, from http://www.nichd.nih.gov/crmc/cdb/r_overview.htm

Mattingly, I. G. (1972). Reading, the linguistic process and linguistic awareness. In J. Kavanaugh & I. Mattingly (Eds.), *Language by ear and by eye* (pp. 133–147). Cambridge, MA: MIT Press.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of scientific research literature on reading and its implications for reading instruction. Reports of the subgroups (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Pearson, D. P., Johnson, D. D., Clymer, T., Indrisano, R., Venezky, R. L., Baumann, J. F., et al. (1991). *World of Reading*. Needham, MA: Silver, Burdett, & Ginn.

Schieffer, C., Marchand-Martella, N. E., Martella, R. C., Simonsen, F. L., & Waldron-Soler, K. M. (2002). An analysis of the *Reading Mastery* program: Effective components and research review. *Journal of Direct Instruction, 2*(2), 87–119.

Science Research Associates (SRA)-McGraw-Hill (1999). *Field testing Horizons: Using the classroom to develop a highly effective reading program*. Columbus, OH: Author.

Shinn, M. R. (Ed.). (1989). *Curriculum based measurement: Assessing special children.* New York: Guilford.

Telian, N. (1997). *Lively letters: Multisensory mnemonic letter card program*. Stoughton, MA: Telian-Cas Learning Concepts.

Torgesen, J. K., Wagner, R. V., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children. *Journal of Educational Psychology, 91,* 579–593.

Vreeland, M., Huth, E., Lum, V., Pattison, R., & Vail, J. (1998). Accelerating cognitive growth: *Horizons Learning to Read Fast Track A–B* program: Research from the field. Unpublished manuscript.

Wilkinson, G. S. (1993). *Wide Range Achievement Test—Revised*. Wilmington, DE: Wide Range, Inc.
Woodcock, R. W. (1997). *Woodcock Diagnostic Reading Battery*. Itasca, IL: Riverside Publishing Company.
Woodcock, R. W. (1998). *Woodcock Reading Mastery Test—Revised*. Circle Pines, MN: American Guidance Service.
Woodcock, R. W., & Johnson, B. M. (1989). *Woodcock-Johnson Psycho-Educational Battery—Revised*. Itasca, IL: Riverside Publishing Company.