

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

The Effects of Beginning Reading Instruction in the Horizons Reading Program on the Reading Skills of Third and Fourth Graders

Abstract: This study is a follow-up on a previous study of the effects of 2 beginning reading programs implemented in 1st grade. In the previous study, 40 1st-grade students who were matched based on their Concepts About Print Test (Clay, 1979) and Phonological Segmentation Fluency from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2003) scores entered 4 different 1st-grade classes. Two classes used *Horizons Fast Track A-B* (Engelmann, Engelmann, & Seitz-Davis, 1997) and 2 classes used the *Silver, Burdett, and Ginn* (Pearson et al., 1991) curriculum. Students in 1 *Horizons* class and 1 *Silver, Burdett, and Ginn* class also received explicit phonological awareness instruction in kindergarten. The current study compares these groups' scores from state mandated tests in 3rd and 4th grades. Students who received reading instruction in *Horizons* had significantly higher scores on the Massachusetts Comprehensive Assessment System in reading at the end of 3rd grade and in English Language Arts at the end of 4th grade. Thus, beginning reading instruction in *Horizons* appears to have broad and long term benefits as compared to *Silver, Burdett, and Ginn*.

Studies sponsored by the National Institute of Child Health and Human Development indicated that 44% of students in fourth grade had poor reading skills (Lyon, 1998). When these children were followed up, their problems were found to have increased because they were unable to read advanced materials such as science and literature. One strategy to combat this national problem is to attempt to remediate the reading skills of these struggling upper elementary students. Another strategy is to optimize reading instruction before poor reading patterns develop. This study examines the role of beginning reading curriculum and phonemic awareness instruction in strengthening the reading skills of third- and fourth-grade students.

Chall (1967) reported the results of 25 reading investigations undertaken between 1900 and 1960. She concluded that focused instruction in phonics was superior to instruction without this focus in teaching word recognition, oral reading, and spelling. These findings held for both low performers and typically achieving students. Adams (1990) reviewed more recent research on beginning reading instruction and came to the same conclusion—that explicit instruction in phonological awareness and code-based phonics programs produced much higher reading achievement than so called meaning-based programs. Recently, Torgesen et al. (1999) found that explicit phonics instruction leads to greater improvement in word reading than typical basal reading programs. Furthermore, Foorman, Francis,

Journal of Direct Instruction, Vol. 4, No. 2, pp. 129–137. Address correspondence to Kevin G. Tobin at ktobin@pittsfield.net

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. Thus, a great deal of research converges on the conclusion that systematic explicit instruction in the phonological structure of the spoken word, followed by systematic phonics, is a critical component of effective early reading instruction.

Direct Instruction is a model of teacher-directed instruction that involves careful analysis of the curriculum, detailed instructional design, teacher modeling, careful prompting, and the immediate correction of errors. Direct Instruction differs from other models of explicit instruction by emphasizing the analysis of curriculum focused on identifying strategic generalizations (big ideas) and the precise design of communications. *Reading Mastery* (Engelmann & Bruner, 1995) is a Direct Instruction reading program that has been developed, evaluated, and researched over a period of 25 years. Schieffer, Marchand-Martella, Martella, Simonsen, and Waldron-Soler (2002) conducted a comprehensive analysis and literature review of the *Reading Mastery* curriculum. Included was a description of how *Reading Mastery* addressed critical areas of reading instruction that have been identified by reading researchers. Next, the authors examined 21 research studies that compared *Reading Mastery* or *Distar Reading* (a previous version of *Reading Mastery*), to other curricula. They found that *Reading Mastery*/*Distar Reading* was favored by a statistically significant margin in 14 (67%) of the studies, and the alternative program was favored in only 3 (14%), with the remaining 4 studies showing no statistically significant differences.

Horizons (Engelmann, Engelmann, & Seitz-Davis, 1997) is a more recently developed Direct Instruction reading program. The *Horizons* reading program has three levels:

Level A, Level B, and Level C–D. The first two levels are also combined into an accelerated version—*Horizons Fast Track A–B*. The *Horizons* reading program incorporates many of the features of *Reading Mastery* but differs in some aspects of how decoding is taught and in the extent of instruction in comprehension strategies. *Reading Mastery* uses a special font (or orthography) as scaffolding for beginning readers; *Horizons* uses a standard font with color and underlining to provide the scaffolding. In both programs, the scaffolding is faded out as students gain skills. In addition, the *Horizons Fast Track A–B* program presents a wider diversity of story reading formats to increase the understanding of characters, plot, and sequence, and allows students to analyze details from the story.

Tobin (2003) found that students who received instruction with *Horizons Fast Track A–B* (with and without prior phonological awareness training) significantly outperformed those students who received the *Silver, Burdett, and Ginn* curriculum (with and without prior phonological awareness training). In this study, 36 first graders were matched for their phonological segmentation skills on the DIBELS in September. Nineteen students in two classes received instruction with the *Horizons Fast Track A–B* reading program, and 19 students in two other classes received instruction in the *Silver, Burdett, and Ginn* curriculum. One class in each group had received prior phonological awareness training in kindergarten. The results revealed that students who had received instruction with the *Horizons* program performed significantly better on DIBELS measures of phonemic segmentation, nonsense word fluency, and oral reading fluency. The *Horizons* students scored significantly higher on the Woodcock Diagnostic Reading Inventory (Woodcock, 1997). The purpose of this study was to evaluate the long-term effect of the *Horizons Fast Track A–B* program on the academic success of students as indicated by state-mandated tests. This study addresses the essential question of whether early reading

success associated with high quality reading instruction is maintained through the third and fourth grades.

Method

The methods regarding selection and assignment of participants, materials, and treatments were described in detail by Tobin (2003). The following section includes brief summaries of these methods. More detail is available in Tobin (2003).

Participants

The study took place in four first grade-classes located in four different schools. Two classes came from schools in more affluent sections of the city, and two classes came from schools with high rates of poverty. One class from a more affluent school and one class from a less affluent school were assigned to each treatment. Table 1 shows the assignment of treatments and the demographics of each of the schools in this study.

Nineteen students received instruction in *Horizons Fast Track A-B*, and 19 students received instruction using *Silver, Burdett, and Ginn*. Students were not randomly assigned; classes were assigned to treatments and students from these classes were selected in order to create maximum similarity across groups. Two measures—The Concepts About Print Test and the Phonological Segmentation Fluency subtest of the DIBELS—were used for matching and to determine initial equivalence.

All students were at least 6 years of age in September of their first-grade year. 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 Concepts About Print and Phonological Segmentation Fluency 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. Twenty female students and 18 males completed the study. Five students were African American and 33 were Caucasian. The proportion of African-American students (13%) to

Table 1
Percent of Free and Reduced Cost Lunch and Racial/Ethnic Diversity by Schools

	Free/ reduced lunch	Ethnic/racial groups			
		White	African American	Hispanic	Asian
<i>Horizons schools</i>					
Class A	44.4%	80.3%	12.5%	3.2%	3.2%
Class B	22.4%	90.9%	5.2%	0.8%	2.7%
<i>Silver, Burdett, and Ginn schools</i>					
Class C	67.0%	62.7%	22.1%	13.3%	1.0%
Class D	17.8%	96.4%	2.5%	0.2%	0.0%

Caucasian students (86%) was approximately the same as in the city where this study took place (88% Caucasian, 9% African American). Tobin (2003) reported that there were no statistically significant pretest differences between groups.

Students in the *Silver, Burdett, and Ginn* condition continued to receive instruction with this curriculum in second grade. Then the school system adopted the Harcourt Brace Reading program and they received instruction in this curriculum for third and fourth grades. Across the district, students are required to receive 90 min of reading instruction per day. Almost all of these students remained in their original schools. One student moved to another elementary school in the district, and one student moved out of the state. These students were not kept together as a group, but were distributed among the classes in their respective schools. Three students in the *Silver, Burdett, and Ginn* condition later became eligible for special education. One student received services for ADHD and reading difficulty, the other two for reading difficulties.

Students in the *Horizons* condition in one school went on to receive instruction in *Silver, Burdett, and Ginn* in second grade. In third and fourth grade, these students received reading instruction in the Harcourt Brace Reading program. One school continued to pilot the *Horizons C-D Fast Track* program. These students received *Horizons* instruction during second grade but then received instruction in the Harcourt program in third and fourth grades. One of the students who received *Horizons* in first grade moved out of the state, and one student moved to another elementary school in the district. The students in the *Horizons* condition were not specially grouped after first grade. They were distributed among the classes in each of their schools. All but two of the *Horizons* students remained in the school where they attended first grade. Two of the students in the *Horizons* condition received special education services. One had received

special education for a physical disability since preschool. The other received resource room services for writing difficulties in third grade.

Measures

Massachusetts Comprehensive Assessment System (MCAS)

The MCAS is the state-mandated accountability test used in Massachusetts. This test is based on the Massachusetts Curriculum Frameworks established in 1993. The MCAS is carefully administered. Detailed administration manuals are given to teachers and principals to ensure that test administration is uniform. The test consists of multiple-choice questions, open-response items, and a writing prompt. At least two professionally trained scorers rate the open-response items and the writing prompt; multiple-choice questions are machine scored. A scoring management system tracks agreement between the two raters. If there is a scoring disagreement, additional raters resolve it. The evidence for the reliability and validity of the MCAS was presented in the 1998 MCAS Technical Report. At the fourth-grade level, 73,527 students took the English Language Arts (ELA) portion of the test, and 74,068 students took the math portion. Cronbach alpha reliability was .90 for ELA and .87 for math. The best evidence for the construct validity of the MCAS is its robust correlation with commercial achievement tests. For example, the ELA portion has a .80 correlation with the Stanford Achievement Test.

The third-grade MCAS test is a reading test. This test consists of eight or nine passages for students to read. Students respond by answering 42 multiple-choice questions and two open-response questions. The open-response questions require students to generate a one or two paragraph response in writing, or in the form of a chart or diagram, as appropriate. The open-response questions are scored using a scoring guide by a trained

rater. For the third-grade reading MCAS there is not an *advanced* category, and scale scores are not reported. On this test the *proficient* range is between 29 and 40, the *needs improvement* range is between 16 and 28, and the *warning* range is below 16.

The fourth-grade MCAS involves two parts. The first part requires students to generate a multiple paragraph essay in response to a writing prompt. This component is administered in two sessions on the same day with a short break between the sessions. In the first session students write a rough draft of their composition in response to a writing prompt. During the second session students revise their work and produce a final draft. Students are allowed to use a dictionary when they create their composition. Two independent raters score student compositions. The students are scored on topic development on a 6-point scale and the use of standard English conventions on a 4-point scale.

The second part of the fourth-grade MCAS consists of a Language and Literature Test. The Language and Literature Test is administered in three separate test sessions. Each session consists of selected readings followed by multiple-choice and open-response questions. No reference materials are allowed in these sessions. Open-response questions are scored using a scoring guide. The scoring guide specifies what knowledge and skills should be demonstrated. On the fourth-grade MCAS tests, a score below 220 is *failing or warning*; a score between 220 and 239 is in the *needs improvement* category. A score between 240 and 259 is *proficient*, and a score between 260 and 280 is *advanced*.

Materials

Horizons Fast Track A–B

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

per lesson that are gradually upgraded in complexity. *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 six parts. First, letters and sounds are practiced. Second, word attack skills are taught. Third, oral reading of a story is conducted. Fourth, story-based activities such as independent workbook activities are conducted. Finally, letter writing, sentence writing, and spelling occur.

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 scripts, 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

Three teachers taught the *Silver, Burdett, and Ginn* curriculum that has been the district's reading curriculum for 15 years. The *Silver, Burdett, and Ginn* program became popular in the mid-1980s. It consists of an anthology of children's literature written by popular children's authors. In subsequent revisions, phonics activities were added; however, these activities were not connected with the stories that were read. Each lesson consists of silent reading and workbook activities. Phonics activities 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 phonemic 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 *Silver, Burdett, and Ginn* 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 *Silver, Burdett, and Ginn* program several years previously by the publisher. *Silver, Burdett, and Ginn* is not a scripted curriculum, and considerable teacher discretion is involved in deciding the pace and order of the lessons.

Results

Table 2 reports descriptive statistics for each classroom as well as for the combined *Horizons* group (classrooms A and B) and the *Silver, Burdett, and Ginn* group (classrooms C and D) on the third-grade Reading MCAS and fourth-grade English Language Arts MCAS. On both tests, students from both *Horizons* classrooms outperformed students from the two *Silver, Burdett, and Ginn* classrooms by substantial margins. The combined *Horizons* groups' mean exceeded the *Silver, Burdett, and Ginn* groups' mean with very large effect sizes, $(M_{Horz} - M_{SBG})/S_{pooled}$, of 1.28 on the third-grade test and 1.56 on the fourth-grade test. On the third-grade test, the median score from the *Horizons* group equaled the highest score from the *Silver, Burdett, and Ginn* group. On the fourth-grade test the median *Horizons* score (243) exceeded 89% (16 of 18) of the scores in the *Silver, Burdett, and Ginn* group. The sta-

Table 2
Means and Standard Deviations of Third-Grade Reading MCAS and Fourth-Grade English Language Arts MCAS

	Third grade			Fourth grade		
	<i>n</i>	Mean	<i>SD</i>	<i>n</i>	Mean	<i>SD</i>
<i>Horizons</i> classes						
Class A	9	36.89	2.76	9	250.67	11.53
Class B	8	32.50	3.25	9	243.11	5.21
Combined	17	34.82	3.68	18	246.89	9.51
<i>Silver, Burdett, and Ginn</i> classes						
Class C	9	31.11	3.52	10	233.40	10.24
Class D	8	26.13	6.83	8	230.00	9.26
Combined	17	28.76	5.76	18	231.89	9.69

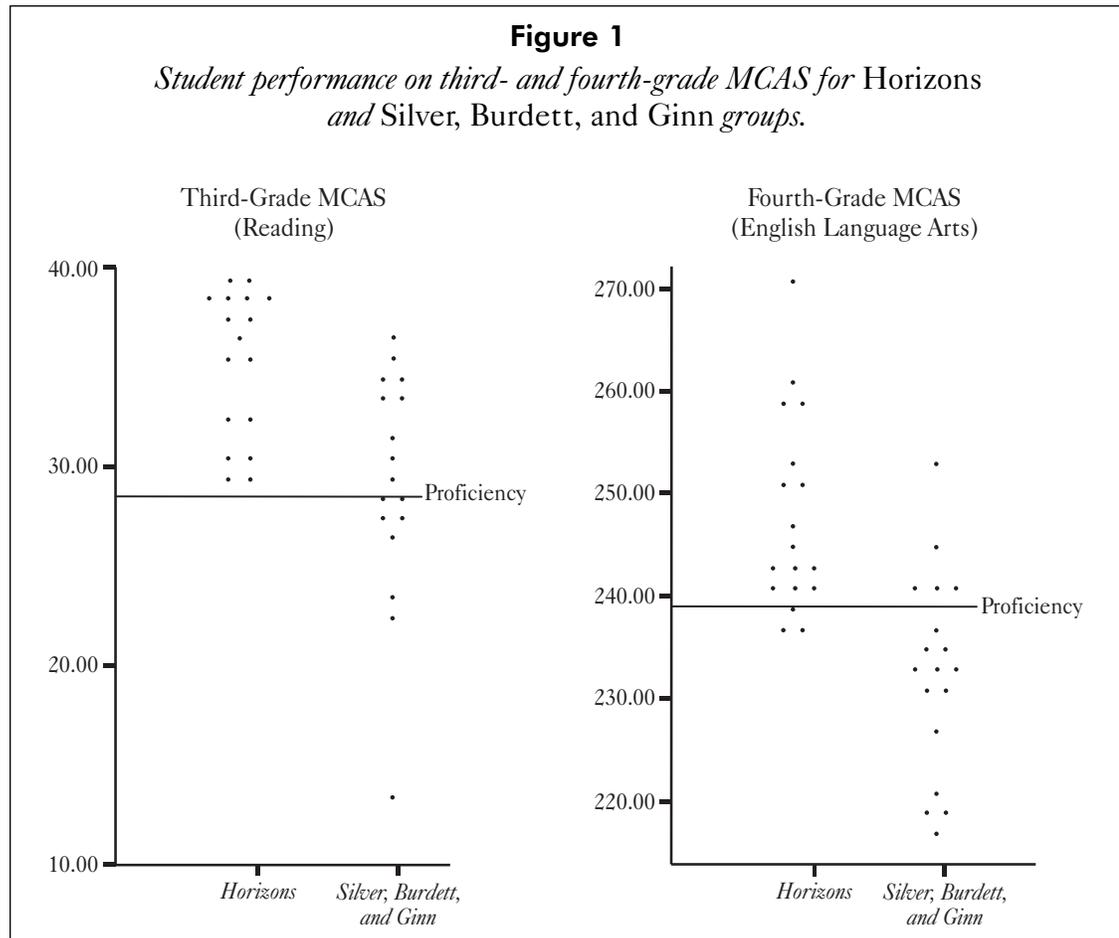
tistical significance of the differences between the *Horizons* and *Silver, Burdett, and Ginn* groups were tested with a one-way ANOVA on each test. The difference was statistically significant on both the third-grade test, $F(1, 32) = 13.36, p = .001, \eta^2 = .29$, and the fourth-grade test, $F(1, 34) = 21.97, p < .001, \eta^2 = .39$.

The differences between groups can be seen clearly in Figure 1. Each dot in this figure represents one student's score. This figure shows that, on both tests, the distribution of scores for the *Horizons* group is shifted upward compared to the *Silver, Burdett, and Ginn* group. The horizontal lines represent the threshold

for proficiency that has been established for the test. On the third-grade test, 100% of the *Horizons* students, but only 53% (9 of 17) of the *Silver, Burdett, and Ginn* students, exceeded this standard. On the fourth-grade test, 83% (15 of 18) of the *Horizons* students scored above the proficient threshold, but only 28% (5 of 18) of the *Silver, Burdett, and Ginn* students achieved this level of performance.

Discussion

The students in this study came to first grade with comparable levels of early reading skill; however, by the end of first grade, stu-



dents who received reading instruction in the *Horizons Fast Track A–B* program significantly outperformed students who received instruction in a conventional reading curriculum on measures of basic reading, word reading efficiency, and general reading skills (Tobin, 2003). Results from this follow-up study indicate that students who received the *Horizons* curriculum maintained and even increased their superior performance 2 and 3 years later on state-mandated tests. On both tests, the differences between the *Horizons* group and the *Silver, Burdett, and Ginn* group were very large.

Limitations of This Study

The results must be considered with some caution. In this study random assignment of teachers and students was not possible. Participating teachers were volunteers, and students were selected based on their assignment to volunteer teachers' classes. Teachers who volunteered to teach *Horizons Fast Track A–B* may have been more venturesome or more engaged than teachers who did not volunteer. Further, the small sample of teachers limits generalizability of these results. Additional studies involving more teachers and different designs are needed before generalization can be made about the relative efficacy of the *Horizons Fast Track A–B* curriculum compared to *Silver, Burdett, and Ginn*. Another limitation of this study was the small number of students and lack of random assignment of students to conditions. The pretest results suggest that groups were comparable prior to the initiation of the treatments; however, additional studies are needed before firm conclusions can be drawn.

Effects of Instruction on State-Mandated Tests

At first glance, it seems implausible that a curriculum intervention conducted in first grade would establish differential trajectories of reading growth that persisted for 3 years. None of the classes remained intact; all of the

students in this follow up had been shuffled into different classes, and in some cases, different schools. This did not seem to matter. The students who received instruction from the *Horizons* program continued to outperform students from a conventional program long after the program ended. This is consistent with the position of Good, Simmons, and Smith (1998). They note that stable reading trajectories can be deduced from the high correlation between reading performance in the early primary grades and later reading skills. Good et al. concluded that it is very difficult for students with low initial skills to make adequate progress and that the optimal solution is to intervene early so that students have sufficient initial skills to continue to make progress. They proposed that the best way to ensure this strong initial start is to provide instruction that involves the development of phonological awareness, systematic alphabetic understanding, explicit and systematic phonological recoding, and instruction focusing on accuracy and fluency while reading connected text. These recommendations correspond very closely with the content of the *Horizons Fast Track A–B* curriculum. This study suggests that giving students a strong initial start can increase success on high stakes state-mandated tests. This has real implications for school districts, especially districts with a high proportion of disadvantaged students.

Massachusetts' regulations require that all students attain a *proficient* score on the MCAS by 2014. Given what is known about reading trajectories, this is likely to be a difficult task unless students start out quite proficient. In the present study 100% of the students who were taught using the *Horizons Fast Track A–B* curriculum scored in the *proficient to advanced* range on the MCAS in third grade and 83% attained this level on the fourth-grade test. These students are well on their way to meeting the state goal. Unfortunately, a substantially lower percentage of students who received early reading instruction in a conventional basal reading program attained a level of

proficiency. It is important to note that half of the *Horizons Fast Track A–B* students worked in this curriculum in first grade only, and half had it for first and second grades. Thus, for many of the students, a single year of intervention in first grade appears to have made an important difference. This supports Good et al.'s (1998) suggestion that first grade is a critical period for reading instruction.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Chall, J. S. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.
- Clay, M. M. (1979). *The Concepts About Print Test*. Exeter, NH: Heinemann Educational Books.
- 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.
- Good, R. H., & Kaminski, R. (2003). *DIBELS: Dynamic Indicators of Basic Early Literacy Skills*. Longmont, CO: Sopris West.
- 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.
- Lyon, G. R. (1998). *Overview of reading and literacy initiatives*. Retrieved May 4, 2000, from the National Institute of Child Health and Human Development Web Site: http://www.nichd.nih.gov/crmc/cdb/r_overview.htm
- Massachusetts Comprehensive Assessment System: MCAS Technical Report Summary. (1999). The Commonwealth of Massachusetts Department of Education. Retrieved September 15, 2003, from http://www.doe.mass.edu/mcas/tech_rpt.html
- Massachusetts Comprehensive Assessment System: MCAS Technical Report. (2002). The Commonwealth of Massachusetts Department of Education. Retrieved September 13, 2003, from http://www.doe.mass.edu/mcas/tech_rpt.html
- Massachusetts Comprehensive Assessment System. (2003). *Grade 4 English Language Arts Test*. The Commonwealth of Massachusetts Department of Education. Retrieved September 15, 2003, from <http://www.doe.mass.edu/mcas/testitems.html>
- 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 a research review. *Journal of Direct Instruction, 2*, 87–119.
- Tobin, K. G. (2003). The effects of the *Horizons* reading program and prior phonological awareness training on the reading skills of first graders. *Journal of Direct Instruction, 3*, 1–16.
- 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.
- Woodcock, R. W. (1997). *Woodcock Diagnostic Reading Battery*. Itasca, IL: Riverside Publishing.