

## The Role of Instruction in Learning to Read: Preventing Reading Failure in At-Risk Children

Barbara R. Foorman  
University of Texas—Houston Medical School

David J. Francis  
University of Houston

Jack M. Fletcher  
University of Texas—Houston Medical School

Christopher Schatschneider  
University of Houston

Paras Mehta  
Arizona State University

First and 2nd graders ( $N = 285$ ) receiving Title 1 services received 1 of 3 kinds of classroom reading programs: direct instruction in letter–sound correspondences practiced in decodable text (direct code); less direct instruction in systematic sound–spelling patterns embedded in connected text (embedded code); and implicit instruction in the alphabetic code while reading connected text (implicit code). Children receiving direct code instruction improved in word reading at a faster rate and had higher word-recognition skills than those receiving implicit code instruction. Effects of instructional group on word recognition were moderated by initial levels of phonological processing and were most apparent in children with poorer initial phonological processing skills. Group differences in reading comprehension paralleled those for word recognition but were less robust. Groups did not differ in spelling achievement or in vocabulary growth. Results show advantages for reading instructional programs that emphasize explicit instruction in the alphabetic principle for at-risk children.

Learning to speak one's native language is a natural process in that explicit teaching is not required. Reading, in contrast, has been called an "unnatural act" (Gough & Hillinger, 1980) to emphasize the fact that one's writing system relates to speech in an arbitrary way and, therefore, has to be taught (Lieberman, Shankweiler, & Lieberman, 1989). What needs to be taught is the alphabetic principle: that letters in a word relate to speech in a conventional and intentional way. For many children, insight into this principle will develop through informal instruction at home and nondirective activities at school. However, as many as one in five children have difficulty learning to read (Lyon, 1995; Shaywitz, Fletcher, & Shaywitz, 1994). There may always be a small

percentage of children who are at risk of reading failure for a variety of cognitive, linguistic, or social–emotional factors. However, in urban settings, there are entire schools in which reading failure is the norm, in part because of lack of home preparation in understanding the alphabetic principle (Adams, 1990) and also because of inadequate instruction in the classroom (Slavin, Karweit, & Wasik, 1994). The importance of learning to read in the early grades is clearly illustrated in a longitudinal study that addressed long-term development of reading skills from kindergarten to Grade 9 (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996). This study showed that, on average, children who were poor readers in Grade 3 did not "catch up" to their peers in their reading skills; the growth of reading skills fit a deficit, not a lag, model. Moreover, 74% of children who were poor readers in Grade 3 were poor readers in Grade 9.

In the last two decades, a scientific body of evidence has accumulated pointing to a phonological processing deficit as the core cause of poor reading (Fletcher et al., 1994; Foorman, Francis, Fletcher, & Lynn, 1996; Lieberman et al., 1989; Stanovich & Siegel, 1994; Wagner, Torgesen, & Rashotte, 1994). Burgeoning evidence exists that deficits in this area can be ameliorated through appropriate training, particularly with younger children in kindergarten through Grade 2 (Ball & Blachman, 1991; Bradley & Bryant, 1983; Foorman, Francis, Shaywitz, Shaywitz, & Fletcher, 1997a; Torgesen, 1997; Vellutino et al., 1996) or as early as preschool (Byrne & Fielding-Barnsley, 1991, 1993, 1995). Ball and Blachman (1991) and Foorman et al. (1997a) supplemented kindergarten programs for children at risk for

---

Barbara R. Foorman and Jack M. Fletcher, Department of Pediatrics, University of Texas—Houston Medical School; David J. Francis and Christopher Schatschneider, Department of Psychology, University of Houston; Paras Mehta, Department of Psychology, Arizona State University.

Portions of this article were presented at the meetings of the American Association for the Advancement of Science, Seattle, Washington, February 18, 1997, and the American Educational Research Association, Chicago, March 25, 1997. This research was supported by National Institute of Child Health and Human Development Grants HD30995 and HD28172.

Correspondence concerning this article should be addressed to Barbara R. Foorman, Center for Academic and Reading Skills, University of Texas—Houston Medical School, 7000 Fannin, UCT #860, Houston, Texas 77030. Electronic mail may be sent to bfoorman@ped1.med.uth.tmc.edu.

reading problems with activities and tasks involving phonological awareness skills. Both studies showed clearly that the supplementation of standard kindergarten curriculums with activities involving phonological awareness skills resulted in growth in phonological awareness skills relative to children who received the standard curriculum without phonological awareness skills. The studies also showed that these gains continued and were also manifested in areas involving word reading in the first and second grades (see Foorman, Francis, Beeler, Winikates, & Fletcher, 1997).

Vellutino et al. (1996) provided either one or two semesters (depending on progress) of 30 min daily, one-on-one tutoring to poor readers in Grade 1. The tutoring in letter identification, phoneme awareness, word-reading skills, and practice in connected text helped the majority of these children become average readers. Torgesen (1997) found that 20 min a day for 80 hr of one-on-one tutoring in phonological decoding strategies (with or without training in articulatory gestures) and practice in reading and writing enabled approximately 75% of first graders who had been in the bottom 10th percentile in phonological skills in kindergarten to move to national averages in timed and untimed decoding. Similar results were achieved with older, severely disabled readers (age 10 years on average); however, the one-on-one tutoring was much more intensive—2 hr daily for 80 hr—and decoding accuracy but not speed reached national averages. Olson, Wise, Ring, and Johnson (1997) had similar results with third to sixth graders below the 10th percentile in word recognition who were tutored individually in phonological decoding strategies (with or without training in articulatory gestures).

The efficacy of the interventions in these studies, which emphasized tutorial interventions, is interesting in relation to older studies that also focus on early intervention. In summarizing these programs, Slavin and his colleagues (Slavin, Karweit, & Madden, 1989; Slavin et al., 1994) noted that the most widely used supplementary-remedial programs, diagnostic-prescriptive pullout programs provided under Title 1 programs for economically disadvantaged children, showed little evidence of effectiveness unless they involved one-on-one tutoring. Moreover, the attempt to mainstream at-risk children by having Title 1 or special education aides work in the regular classroom has been no more effective than the pullout model (Archambault, 1989; Puma, Jones, Rock, & Fernandez, 1993).

In contrast, kindergarten or first-grade prevention programs and classroom change models have proved effective. The only prevention programs for which data are available on long-term effects of intensive reading instruction in the early grades are Reading Recovery (Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994; Shanahan & Barr, 1995) and Success for All (Slavin, Madden, Dolan & Wasik, 1996). In evaluations of Reading Recovery, first graders tutored daily for 30 min by a trained Reading Recovery tutor exceeded matched control children's reading performance with an effect size of .87. This effect size fell to .45 and .29 one and two years later, respectively, without additional intervention. More recent analysis of the effects of Reading Recovery continue to show large effect sizes that diminish over time. Reading

Recovery can more quickly recover children to middle reading group levels if it is modified to include direct instruction in the alphabetic code (Iverson & Tunmer, 1993), and other programs may provide equally large effects without the tutorial component (Shanahan & Barr, 1995).

Classroom change models are based on the assumption that the best way to minimize the need for remedial services is to provide the best possible classroom instruction in the first place. A more traditional kind of classroom change model is what Slavin et al. (1989) referred to as "continuous progress models." Students in these classrooms proceed at their own pace through a sequence of well-defined instructional objectives. They are taught in small groups on the basis of skill level and are frequently assessed and regrouped on the basis of these assessments. The best known of these programs is DISTAR (Engelmann & Bruner, 1995; now SRA Reading Mastery), a highly structured and scripted program that has produced positive results in many large-scale studies (see Aukerman, 1984; Shanahan & Barr, 1995).

Although programs such as Reading Recovery, SRA Reading Mastery, and Success for All show good efficacy, they have not attempted to isolate the components of effective reading instruction. Current research suggests that a necessary skill to be mastered in learning to read in the early grades is decoding. Decoding typically refers to the application of the letter-sound correspondences taught in phonics. Although decoding is more accurately described as deciphering the printed word, and phonic rules may simply play an attentional role in the weightings of connections between orthographic and phonological units (Adams, 1990; Foorman, 1994), decoding accuracy is the single best predictor of reading comprehension (Stanovich, 1990; Vellutino, 1991). Thus, an instructional focus on developing decoding skills early in school is consistent with the relationship of decoding skills and comprehension, especially for children whose only chance to learn to read is in school. An important question is how explicit decoding instruction needs to be, whether highly explicit through decontextualized letter-sound correspondence rules practiced in controlled vocabulary text or implicit through incidental learning gained by feedback on reading literature. "The Great Debate" over code-emphasis versus meaning-emphasis approaches to reading captures the extremes of this continuum of explicitness (Chall, 1983; Foorman, 1995a, 1995b). However, there is the middle ground of embedded-phonics approaches in which instruction in letter-sounds and spelling patterns is contextualized within literature selections.

In the present article, we investigated questions involving the degree of explicitness in alphabetic code instruction and effects of phonological processing on growth in word reading in children at risk for reading failure traditionally served in Title 1 programs. In a large sample of children receiving Title 1 services, we hypothesized that children who received explicit instruction in the alphabetic principle with an emphasis on letter-sound correspondences would show greater growth over 1 school year of classroom instruction relative to children receiving less explicit instruction focusing on spelling patterns or children receiving

implicit instruction in the alphabetic principle. We also hypothesized that this growth in reading skills would be moderated by initial phonological processing skills.

### Method

#### Participants

Participants were 285 of the 375 children in first and second grades eligible for services under Title 1 funding in an urban district with 19 elementary schools. The 90 children were excluded from the present analyses because they had been placed on a wait list and never did receive Title 1 services during the study. Thus, analyses are restricted to those eligible students who actually received tutoring during the year.

Title 1 refers to federal funding provided for economically disadvantaged children with low achievement. Economic disadvantage is usually defined in terms of the percentage of children participating in the federal lunch program, as it was in this study. Low achievement was defined by school district officials as scores on the district's emergent literacy survey in the bottom quartile in first- and second-grade classrooms at each Title 1 school. Hence, although all children in the lowest quartile received the classroom interventions, the present sample represented the lowest 18% because of lack of funds for tutoring.

The participating children attended 8 of the 10 Title 1-eligible elementary schools in this district. (The Title 1 program was in its 2nd year of implementation in the district.) The percentage participation in the federal lunch program ranged from 32.3% to 71.4% at the 8 schools. Thus, the participating children were only those 3 to 8 children in each regular education classroom who were served through Title 1 in the participating schools. The non-Title 1 children in the classrooms were not participants in the study, at the request of district officials; however, they received the same classroom curricula as the participating children.

School participation was determined by the willingness of the principal and teachers to participate. The design called for some schools to have only one instructional approach and for others to have two approaches in an attempt to control for school effects. The design is described in Table 1, which provides information on the number of classrooms per grade receiving each of the four curricula. No second-grade classrooms are listed for Schools 4 and 5 because Title 1 funds were available only to serve first graders. Also, it is important to note that the school selected by district officials to be the unseen comparison had the largest total enrollment, the largest percentage of children participating in the federal lunch program (71.4%), and the lowest achievement scores on the statewide test in Grade 3. To deal with what was widely perceived as a "tough" school, district officials placed a well-respected principal and Title 1 teachers at the school; nonetheless, the school was not regarded as a desirable teaching assignment by classroom teachers.

The ethnic composition of the sample was as follows: 60% African American, 20% Hispanic, and 20% White. The ethnic composition of the district at large was approximately 20% Asian, 26% African American, 23% Hispanic, and 31% White. Sixty-one percent of the sample was male. Instructional groups did not differ in age, gender, or ethnicity.

#### Instructional Methods

During the 90-min daily language arts period, the children were instructed in one of three classroom reading methods, all of which existed within a literature-rich environment in the classroom: direct

Table 1  
*Study Design and School Characteristics*

School	Enrollment	Federal lunch program (%)	Grade	No. classrooms	Curriculum
1	1,208	71.4	1	5	IC-S
			2	5	IC-S
2	1,009	49.5	1	6	IC-R
			2	4	IC-R
3	1,232	64.2	1	6	EC
			2	6	IC-R
4	908	43.2	1	3	DC
5	887	41.8	1	2	DC
6	1,137	39.9	1	2	IC-R
			1	2	DC
			2	2	DC
			2	2	IC-R
			2	3	IC-S
			2	2	EC
7	853	64.5	1	2	EC
			1	2	DC
			2	2	EC
			2	2	DC
8	839	32.3	1	3	IC-R
			1	3	EC
			2	2	IC-R
			2	1	EC

Note. IC-S = implicit code-standard; IC-R = implicit code-research; EC = embedded code; DC = direct code.

instruction in letter-sound correspondences practiced in decodable text (direct code [DC]); less direct instruction in systematic spelling patterns (onset rimes) embedded in connected text (embedded code [EC]); and indirect, incidental instruction in the alphabetic code embedded in connected text (implicit code [IC]). The IC condition was either the district standard curriculum (IC-S) or a research implementation developed to ensure comparability of training across instructional approaches (IC-R). Each condition was directed by an advanced graduate student who had been a teacher and who had expertise in professional development, and did not include the authors of this study.

In DC the emphasis was on a balance of phonemic awareness, phonics (with blending as the key strategy), and literature activities, using Open Court Reading's (1995) *Collections for Young Scholars*. Phonemic awareness activities dominate the first 30 lessons of Open Court. The 42 phonic rules are introduced in Lessons 11 through 100, using sound-spelling cards, alliterative stories, and controlled vocabulary text that practice the rule just taught. At the same time decodable texts are used, a parallel strand of Big Book reading occurs so that skills in oral language comprehension and love of story can be developed. Spelling dictation exercises move students from phonetic spellings toward conventional spelling based on phonics knowledge and spelling conventions. Writing workshop activities and anthologies of fiction, nonfiction, and poetry are introduced by mid Grade 1.

In EC the emphasis was on phonemic awareness and spelling patterns in predictable books, using an adaptation of Hiebert, Colt, Catto, and Gary's (1992) program. Teachers providing EC instruction used a common list of sequenced spelling patterns and a guide prepared by participating teachers that listed library books that contained the spelling patterns (see Appendix A for the list of spelling patterns). Whole-class activities such as shared writing, shared reading, choral or echo reading, and guided reading

provided the context for EC instruction. In addition to a general emphasis on a variety of comprehension strategies, EC teachers used the following format in providing strategic guidance about patterns of words: Initially, the teacher would frame a word containing the target spelling pattern during a literacy activity (e.g., *bat*). By deleting the initial phoneme (e.g., *b*), the pattern would be extracted from the word (e.g., *at*). By substituting alternative beginning sounds, students could extend the pattern to new words (e.g., *mat*, *cat*, *hat*). Then students were to identify the target pattern as they encountered it in additional shared and independent reading and writing activities. Finally, patterns were reviewed in the context of reading and writing activities and were incorporated into spelling lists. When the children were working in small groups, they were able to practice these "make-and-break" activities with magnetic letters and acetate boards, always writing down their constructed words and reading their written constructions back to the teacher.

At the time of this study, the staff development in this school district emphasized an IC approach to reading instruction. Central to this IC approach was the emphasis on a print-rich environment with the following characteristics: teacher as facilitator rather than director of learning; children's construction of meaning as central; the integration of reading, spelling, and writing into literary activities that provide a context for phonics; emphasis on classroom interaction and on response to literature; learning centers; and assessment based on portfolios rather than norm-referenced tests (see Routman, 1991; Weaver, 1994). The 19 teachers who participated in the research version of IC worked with the project director—an experienced doctoral-level teacher-trainer who espoused whole-language methods—to define the whole-language philosophy behind their approach:

Whole language is a child centered philosophy of learning and instruction, the implementation of which results in a risk-free, supportive, language-rich environment. This environment is ever-changing; changing to meet the needs of all participants, teachers and students alike. Within this whole language philosophy, students are given a wide variety of opportunities to read, write, learn, and construct meaning within a meaningful context. In this interactive, student-friendly learning atmosphere, learning is not only active and meaningful, but also fun, with the ultimate goal being to instill the desire for life-long learning.

Because of the IC belief in children as readers and writers, even at this "emergent" phase of first and second grades, the emphasis was on learning to foster a competence rather than on learning to perform a skill (see, e.g., Dahl & Freppon, 1995). The use of predictable books and emphasis on writing in this IC approach appear similar to those in the EC approach described previously. However, in the EC approach, the teachers used a systematic list of spelling patterns to teach an analogy strategy for decoding words. In the IC approach, in contrast, the teacher used shared- and guided-reading activities to draw children's attention to specific words or word forms, letters, sounds, patterns, meanings, making predictions, listening for rhymes, and exploring the use of strategies, grammar, language use, spellings, or key ideas in the text. Thus, the opportunity to learn the alphabetic code was incidental to the act of making meaning from print.

In this study, there were 19 IC-R teachers, 20 EC teachers, 14 DC teachers, and 13 IC-S teachers, all of whom volunteered to participate. The IC-S teachers delivered the district's standard instructional method and were trained and supervised by district personnel. Teachers delivering IC-R, EC, and DC were trained during 1 week of summer in-service (30 hr) followed by retraining and demonstration lessons 1 month into the school year. Training

was conducted by members of the research staff, all of whom had previous elementary school teaching experience and were strong proponents of the approach for which they were responsible. During summer in-service, the staff members provided background for the research, discussed instructional strategies relevant to their approach, and worked with teachers to develop a monitoring checklist of the components of the curriculum being implemented. To ensure adequacy of monitoring and control of time on task, all primary reading instruction occurred in 30-min blocks as part of the 90-min language arts block mandated by the state. Because DC used basal materials that were new to the teachers, a representative from the publisher spent 1 day orienting the teachers to the materials. The EC materials were also new, but the project director for this component had considerable experience with onset-rime approaches. During the school year, the research staff visited each teacher's classroom every other week or more frequently, if necessary, to monitor implementation of instruction and to provide feedback on the quality of implementation. Instructional supervisors from the district were available at each school to help teachers with basic issues of classroom management, a resource that was called on infrequently. Research staff members met with the teachers of a particular grade level at each school during their planning time to discuss instructional issues. Finally, to share instructional strategies across sites, teachers implementing a common program in different schools came together after school three times during the school year.

In addition to these 66 classroom teachers, 28 Title 1 teachers delivered one-to-one or small-group tutorials with 3 to 5 students for 30 min each day. In these tutorials, the instructional method either matched that of the classroom or was the district's standard tutorial based on Clay's (1991) method. Because the standard tutorial was an IC approach, there was no mismatch condition for children in the IC-S and IC-R groups.

### Measures and Procedures

*Teacher compliance and attitudes.* During summer training, the teachers in each instructional group and the research staff developed a list of instructional components to be used for bimonthly monitoring of instruction (see Appendix B for the list of each instructional group). The teachers agreed that the monitoring would take place during the 30-min section of the 90-min language arts block, when the focus would be on the reading lesson (which addressed at least the first four components of each instructional approach listed in Appendix B). Occasional visits were made during other times in the language arts block to see how writing and spelling activities progressed and, in the case of the IC-R group, were integrated with reading.

In addition to the checklist used for monitoring, lesson plans were copied, kept, and reviewed as part of compliance. For the monitoring checklist, independent raters were used, with extremely high interrater reliability ( $\geq .80$  for all raters). At the end of the year, we asked the teachers to respond to five questions about their instructional program (see Appendix C for the actual questions). Using a scale ranging from 1 (*definitely yes*) to 5 (*definitely no*), teachers responded to the first four questions asking whether they would recommend the continued use of this approach to instruction. The fifth question asked about the match between the instructional approach delivered and the teacher's beliefs about how to teach children to read; response options ranged from *an exact match* to *not similar at all*.

*Measures given to estimate growth.* Changes in vocabulary, phonological processing, and word-reading skills were assessed four times during the year, in October, December, February, and April. To assess growth in receptive vocabulary, we administered

the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981) four times a year. Both forms (L and M) were used and were alternated in two different sequences. To assess changes in reading skills over the course of the intervention, we asked the children individually to read 50 words aloud that were presented one at a time on 4 × 6-in. cards. The words were matched for frequency of occurrence (Carroll, Davies, & Richman, 1971), were representative of a diversity of linguistic features, and spanned first- through third-grade level of difficulty. Scores were based on the number of words read aloud correctly out of 50. Reliability for the word list was excellent (internal consistency estimate of .9). Concurrent and predictive validities for the word list were also high, as evidenced by correlations exceeding .8 with the Letter Word and Word Attack subtests of the Woodcock-Johnson Psycho-educational Battery-Revised (WJ-R; Woodcock & Johnson, 1989) collected at the end of the year in our normative sample (Foorman et al., 1996).

Phonological processing was measured by the synthesis and analysis tests in the Torgesen-Wagner battery (Wagner, Torgesen, & Rashotte, 1994; see also Foorman et al., 1996, 1997b). The synthesis tests consisted of blending onset rime (*m-ouse*), blending phonemes in real words (*f-a-t*), and blending phonemes in non-words (*m-i-b*). The analysis tests consisted of (a) first sound comparison (in which children were asked to point to the one picture of three that started with the same sound as a target picture); (b) elision (dropping the initial, final, or middle sound of a spoken word); (c) sound categorization (naming the nonrhyming word from a set of four spoken words); and (d) segmentation of a spoken word into phonemes. Each test consisted of demonstration items and 15 test items. In this report we used estimated factor scores that ranged continuously from 0 to 4. Factor score weights were derived from data on a large normative sample from the same school district (Foorman et al., 1996).

*End-of-year achievement and intellectual tests.* At the end of the year, we individually administered the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974) and standardized reading and spelling tests. For the reading tests, we used the WJ-R (Woodcock & Johnson, 1989) to measure decoding (using the Letter-Word Identification and Word Attack subtests) and reading comprehension (using the Passage Comprehension subtest). We used the Formal Reading Inventory (FRI; Wiederholt, 1986) to measure comprehension of narrative and expository text. For spelling we used the Spelling Dictation subtest from the Kaufman Test of Educational Achievement (KTEA; Kaufman & Kaufman, 1985). We did not administer a standardized reading test at the beginning of the year because tests such as the WJ-R lack a sufficient number of items to discriminate initial reading levels for beginning readers and are not adequately sensitive to change over short time intervals.

*Attitude-experience.* In addition to these measures of growth in cognitive skills, academic outcomes, and intellectual abilities, we also collected school attendance data and measures of self-esteem, reading attitudes and experience, behavior, and environmental information in the spring. We assessed self-esteem with a pictorial version of Harter's (1982) Perceived Competence Scale (Harter & Pike, 1984). The five domains of self-esteem assessed were scholastic competence, athletic competence, social acceptance, physical appearance, and behavior or conduct. Children's attitude toward reading was assessed with 11 questions about the extent to which the child enjoyed reading (drawn from the work of Juel, 1988) and 8 questions about whether the child engaged in a variety of literacy experiences. Both the Harter scales and this reading attitude-experience measure use a structure alternative format to minimize the likelihood of the child making the socially desirable response. For each item, children first decide whether the

statement is true or not true about themselves and then decide whether the statement is sort of true or very true. For example, the first item on the reading attitudes measure is "This child [pointing to figure on examiner's left] likes people to read to him/her. This child [pointing to figure on examiner's right] doesn't like people to read to him/her. Which child is most like you? [Child chooses.] Is this child a lot like you or just sort of like you?" Orientation of positive and negative stems of questions and accompanying stick figures varies randomly across items. Items on both the Harter and the reading attitude measures are scored from 1 to 4.

*Teacher evaluations.* The Multigrade Inventory for Teachers (MIT; Agronin, Holahan, Shaywitz, & Shaywitz, 1992) provided a mechanism for the child's classroom teacher to record observations on a rating scale that includes precise descriptions of a full range of behavioral styles reflecting the child's processing capabilities, adaptability, behavior, language, fine motor, and academic proficiency. At the same time, the teacher is able to provide an overall impression of that child's academic strengths and weaknesses and also indicate concerns. The MIT includes 60 items coded by the teacher on a scale ranging from 0 (*never*) to 4 (*often*). There are six scales: Academic, Activity, Language, Dexterity, Behavior, and Attention.

The teacher also completed an end-of-year evaluation, recording the results of pupil placement team meetings and indicating any special services received by the child, recommendations for the next class placement, and recommendations for special services. Grades, absences, tardiness, and results of hearing and visual screening were also recorded. The teacher identified children thought to have emotional, behavioral, or family problems.

## Analysis

We used individual growth curves methodology to analyze changes in phonological processing, word reading, and vocabulary. These methods permit the estimation of (a) the mean rate of change and an estimate of the extent to which the individual's growth differs from this mean rate, and (b) correlates of change, which in this investigation focused on effects resulting from the four instructional groups but also included covariates of verbal IQ, age, and ethnicity. In the analysis of growth in word reading, we also examined the effects of initial level of phonological processing as a correlate of growth and a moderator of instructional effects. Individual growth parameters and correlates of change were estimated using Hierarchical Linear Models-3 (HLM-3; Bryk & Raudenbush, 1987, 1992; see Francis, Fletcher, Stuebing, Davidson, & Thompson, 1991; Francis et al., 1996; Rogosa, Brandt, & Zimowski, 1982, for information on the application of individual growth models in psychology and education). In addition to time being nested within individuals, students were nested within teacher, providing for a three-level model (time, student, teacher). Although teachers are also nested within school, there was an insufficient number of schools to model school-level variability, so this factor was ignored in the analyses.

In analyzing instructional effects, we were first interested in knowing whether IC-R (representing research-trained and monitored instruction) differed from the district's standard (representing district-trained and supervised instruction), tested at  $p < .05$ . Then, to control for Type I error, we conducted Bonferroni-adjusted pairwise comparisons among the three experimental approaches to instruction with an alpha level of .0167 (or .05/3). In modeling academic outcomes, we have ignored differences between IC-S and DC and between IC-S and EC, because these curricula differ from IC-S both in the explicitness of code instruction and in the training of teachers to deliver the instruction. Comparison of IC-S to IC-R provides information about the importance of the teacher-training

component of the study, whereas comparisons among IC-R, DC, and EC provide the critical information about instructional differences controlling for teacher training. In modeling changes over time, we centered age around the last occasion of measurement for each child so that the intercept represented *expected* performance in April. Because we expected older children to outperform younger children, age differences between children at the final assessment were measured as deviations from mean age and were used to predict expected performance and change in performance.

To characterize the pattern of change over time, we fit models to determine (a) whether growth was linear or curvilinear and (b) which of the growth parameters varied across children. This process involved fitting at least the following models: (a) straight line growth with random intercepts and fixed slopes; (b) straight line growth with random intercepts and slopes; (c) curvilinear growth with random intercepts and fixed slopes and quadratic terms; (d) curvilinear growth with random intercepts and slopes and fixed quadratic terms; and (e) curvilinear growth with random intercepts, slopes, and quadratic terms. In all models, errors are assumed to be independently and normally distributed with equal variance over time. A fixed parameter has a value that does not vary across participants, whereas a random parameter has a value that differs across participants. If the mean value for a parameter was not different from zero, and there was no evidence that the parameter differed across participants, then the parameter was dropped from the model. Growth curve analyses for reading, vocabulary, and phonological processing showed that change could be best modeled with linear and quadratic effects and random slopes and intercepts.

## Results

### Tutoring Effects

We examined the size of the tutoring unit (one-to-one or small group, i.e., 3–5 students with one teacher) and the nature of the content of the tutorial (whether it matched or did not match classroom instruction). The mismatch condition was available only for the two code-emphasis groups because the district's standard tutorial—Reading Empowerment, based on Clay's (1991) method—was matched with the IC approach. Unfortunately, it was impossible to retain the initial assignment to ratios of one-to-one or one-to-many because the teachers needed to rearrange groupings to deal with behavioral and learning problems. Thus, we calculated the average number of days a student was in a 1:1 or 1:many ratio condition. This variable did not significantly predict reading growth or outcomes. There was also no significant effect of matched or mismatched tutorial content. Because of the lack of tutoring effects, tutoring was ignored in subsequent analyses.

### Compliance and Attitudes

Compliance data consisted of each teacher's total percentage of compliance in delivering the instructional practices appropriate to her instructional group, as determined from the research staff's monitoring data. Among the 53 classroom teachers monitored (excluding the 13 IC-S teachers, who were not monitored), compliance was generally very high, a median of 80%, with a significant negative skew to the distribution of scores. Four teachers had 0% compliance:

2 were in IC-R, 1 in DC, and 1 in EC. In all four cases, the teachers were teaching reading but were not using the research approach for which they had been trained. The DC and EC teachers were doing the district standard IC-S, as they had been doing for years. The two IC-R teachers were decontextualizing phonics and spelling instruction with work sheets they had purchased. Attempts to retrain and redirect these four teachers met with repeated resistance. We retained these teachers and their students' data in our analyses because they are representative of the range of teaching behaviors encountered in a study of this sort. In short, compliance of 49 of 53 classroom teachers was excellent.

In addition to high compliance with instructional practice, teachers also had positive attitudes toward their instructional method. The distribution of responses for the teacher attitude data for 48 of the 53 research-trained teachers are presented in Table 2 (2 DC, 2 EC, and 1 IC-R teachers did not return the survey). Analysis of variance (ANOVA) using the Kruskal-Wallis test showed significant instructional group differences on the following two questions: "If you were responsible for curriculum decisions in your district, would you recommend that resources (materials, staff development, etc.) be provided for this instructional approach in the future?",  $F(2, 44) = 3.58, p = .036$ ; and "Would you recommend the instructional approach you are using to a colleague?",  $F(2, 44) = 5.23, p = .009$ . Pairwise contrasts

Table 2  
Frequency Distributions for Teacher Attitude  
Survey Data (%)

Question	Frequency distributions				
	Definitely yes 1	Endorse 2 3 4			Definitely no 5
1. Recommend to district					
DC	64	36	—	—	—
EC	22	50	17	11	—
IC-R	44	39	17	—	—
2. Recommend to colleague					
DC	73	27	—	—	—
EC	22	50	11	17	—
IC-R	28	50	17	—	—
3. Recommend for all children					
DC	55	27	18	—	—
EC	28	39	17	11	6
IC-R	33	39	28	—	—
4. Recommend for special needs					
DC	45	27	27	—	—
EC	50	28	11	11	—
IC-R	17	44	28	11	—
	Exactly match	Very similar	Somewhat similar	Not similar at all	
5. Matches my beliefs					
DC	9	82	9	0	
EC	—	61	39	0	
IC-R	22	50	28	0	

Note. DC = direct code; EC = embedded code; IC-R = implicit code-research.

revealed that DC teachers were more likely than EC teachers to recommend their instruction to the district,  $F(1, 44) = 6.95, p < .012$ . Additionally, DC teachers were more likely than either EC or IC-R teachers to recommend their instruction to a colleague,  $F(1, 44) = 9.71, p < .003$  and  $F(1, 44) = 6.80, p = .012$ , respectively. Teachers in the DC, EC, and IC-R groups did not differ in their attitude about recommending their approaches for all children or for children with special needs or in the degree to which the instruction they delivered matched their beliefs about how to teach children to read.

### *Analyses of Baseline Differences in October*

Means and standard deviations for phonological processing and word-reading scores at each wave of data collection are presented in Tables 3 and 4, respectively, for each instructional group according to grade. Correlations between phonological analysis and synthesis factors were greater than .9 at each of the four time points. Therefore, we have elected to present only the results for phonological analysis here (subsequently to be referred to as phonological processing). ANOVA on October baseline scores in word reading and in phonological processing (with age as a covariate) showed no significant differences between instructional groups,  $F(3, 272) = .33, p = .81$ , for word reading; and  $F(3, 271) = 1.87, p = .14$ , for phonological processing.

### *Growth Curve Analyses*

The second graders had minimal reading skills, necessitating the use of first-grade instructional materials with them. Because all children were receiving the same grade-level curriculum, analyses were conducted with age rather than grade as a factor. Exploratory analyses showed that there was no remaining variability in outcomes resulting from grade once age effects were controlled.

Growth curve analyses were conducted using a three-level model: time within child within classroom. All growth curve analyses were conducted using HLM-3 software (Bryk & Raudenbush, 1992). HLM-3 reports tests of fixed effects

using a  $t$  statistic and  $p$  value derived from the unit normal distribution. As a measure of the effect of the instructional group variable, we report  $\Delta R^2$ , which is the proportion of true, between-teacher variance (Level 3) in a growth parameter that is accounted for by the instructional group variable after controlling for all covariates (Bryk & Raudenbush, 1987; Francis et al., 1991). This measure indicates how much of the true, between-teacher variance in slopes and intercepts is uniquely attributable to the instructional methods employed by the teachers. In addition, Cohen's standardized effect size,  $f$  (Maxwell & Delaney, 1990), was computed for curriculum effects as follows. For overall effects of the instructional group variable, we computed the effect ( $\alpha_j$ ) for each group, where  $\alpha_j$  is the difference between the mean value of a parameter (e.g., slope or intercept) in that instructional group and the overall grand mean value for that parameter, taking into account all covariates. The average squared effect was then expressed relative to the HLM-3 estimated error variability in that parameter. This estimate is not printed directly by HLM-3 but can be computed from HLM-3's estimate of the reliability of the parameter and of the systematic variance in the parameter. To estimate the error variance in the instructional group mean growth parameters, we calculated  $[(1 - R)T]/R$ , where  $R$  is the estimated reliability of the random parameter and  $T$  is the estimated systematic variability in the parameter. These two estimates were taken from the growth curve models that included all covariates but did not include the instructional group variable. The square root of this ratio (average squared effect/error variance) gives the standardized effect size,  $f$ . Effect sizes are also reported for differences in growth parameters between specific curricula. These were computed by taking the mean parameter difference between the two curricula and dividing by the square root of the error variability, as just described. Effect sizes for end-of-year outcomes were derived from SAS PROC MIXED (SAS Institute, 1997) two-level random-effects models using a similar approach. However, in these cases, error variability was estimated as the residual variance in an unconditional model divided by the average sample size per classroom.

Table 3  
*Factor Score Means, Standard Deviations, and Sample Sizes for Phonological Processing at Each Wave of Data Collection*

Instructional group	October			December			February			April		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Direct code												
Grade 1	0.68	0.54	44	1.34	0.69	42	1.87	0.74	39	2.16	0.83	41
Grade 2	1.74	0.80	14	2.06	0.47	14	2.25	0.69	14	2.51	0.60	14
Embedded code												
Grade 1	0.37	0.36	49	0.72	0.60	46	1.07	0.69	41	1.59	0.77	39
Grade 2	1.38	0.74	36	1.61	0.62	35	1.89	0.71	29	2.18	0.71	28
Implicit code-research												
Grade 1	0.51	0.55	57	0.93	0.74	57	1.23	0.87	55	1.53	0.88	53
Grade 2	1.58	0.62	28	1.89	0.72	28	2.17	0.79	27	2.21	0.73	25
Implicit code-standard												
Grade 1	0.43	0.50	24	0.90	0.84	24	1.02	0.75	23	1.22	0.86	23
Grade 2	1.48	0.70	24	1.76	0.79	24	1.72	0.63	23	1.90	0.64	22

Table 4  
Raw Score Means, Standard Deviation, and Sample Sizes for Word Reading at Each Wave of Data Collection

Instructional group	October			December			February			April		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Direct code												
Grade 1	0.20	0.51	44	2.17	2.95	42	6.44	7.13	39	12.68	10.21	41
Grade 2	5.73	6.66	15	8.57	7.69	14	12.71	9.60	14	19.43	10.03	14
Embedded code												
Grade 1	0.18	0.88	49	0.72	1.61	46	1.90	2.77	41	5.00	8.15	39
Grade 2	4.75	4.92	36	7.46	6.77	35	12.86	11.04	29	18.29	12.02	28
Implicit code-research												
Grade 1	0.07	0.32	57	0.57	1.20	58	1.20	2.30	55	5.23	7.20	53
Grade 2	5.12	5.24	28	7.96	6.97	28	10.93	9.83	38	16.16	14.32	25
Implicit code-standard												
Grade 1	0.13	0.61	24	0.21	1.02	24	0.57	1.59	23	1.91	2.81	23
Grade 2	3.17	4.90	24	5.36	7.31	24	9.13	7.87	23	14.27	9.35	22

*Analysis of growth in phonological processing.* In the analysis of phonological processing, there were significant differences between ethnic groups and individual differences in age and verbal IQ. African American children had significantly lower expected scores in April than the sample average ( $t = 2.90, p = .004$ ) but did not differ in slope or in the quadratic trend ( $p > .05$ ). Age at the final assessment was a significant predictor of expected score in April ( $t = 4.75, p < .001$ ) and slope ( $t = 3.01, p = .003$ ). This means that older children had higher April scores but improved at a slower rate compared with younger children. Verbal IQ was a significant predictor of expected score in April, slope, and the quadratic effect ( $t = 6.86, p < .001$ ;  $t = 2.81, p = .005$ ; and  $t = 4.05, p < .001$ , respectively). Thus, higher IQ children tended to have higher phonological processing scores in April, but their rate of learning tended to taper off in the latter part of the school year.

There were significant differences in growth in phonological processing among the four instructional groups, controlling for ethnicity and for individual differences in age and verbal IQ. The overall effect of instructional group was large on both intercepts ( $\Delta R^2 = .88, f = 0.69$ ) and slopes ( $\Delta R^2 = .86, f = 1.13$ ). More specifically, children receiving DC had significantly higher scores in April than EC students ( $t = 2.99, p < .003, f = 1.06$ ), and students receiving IC-R ( $t = 4.58, p < .001, f = 1.61$ ). Instructional groups differed significantly in their learning curves. These differences are shown in Figure 1 both for raw scores in the top panel (i.e., observed data) and predicted scores in the bottom panel (i.e., estimates based on the fitted growth model). As is apparent from the predicted scores (panel b), the rate of change in phonological processing scores for the EC group differed significantly from that of the IC-R group and DC groups ( $t = 3.35, p = .001, f = 2.64$ , and  $t = 1.99, p = .045, f = 1.06$ , respectively), although the EC-DC difference is not significant at the Bonferroni-adjusted critical value. In general, the EC group was characterized by a relatively constant rate of change, whereas the IC-R group showed a slowing of growth at the end of the year.

*Analysis of growth in word reading.* Growth in word reading was best described by a quadratic model. In the conditional models, there were no significant effects of ethnicity ( $p > .05$ ), and the effects of age and verbal IQ were similar to those found for phonological processing. Specifically, age at last assessment was a significant predictor of expected performance in April (i.e., the intercept) ( $t = 4.41, p < .001$ ) and the rate of change (i.e., slope;  $t = 2.49, p = .013$ ). Verbal IQ was also a significant predictor of intercept and slope ( $t = 3.70, p < .001$  and  $t = 4.15, p < .001$ ).

Differences between the IC-R and IC-S groups on April performance ( $p > .05, f = 0.16$ ) and growth in word reading ( $p > .05, f = 0.01$ ) were neither statistically nor practically significant. However, there were clear differences among the instructional groups (overall  $\Delta R^2 = .35, f = 0.46$  for intercepts and  $\Delta R^2 = .54, f = 0.24$  for slopes). Controlling for individual differences in age and verbal IQ as well as for ethnicity, DC children improved in word reading at a faster rate than IC-R children ( $t = 2.80, p = .006, f = 0.58$ ) and EC children ( $t = 2.25, p = .024, f = 0.46$ ), although the DC-EC difference is not significant at the Bonferroni-adjusted criterion. Relative to the DC group, the IC-R group's rate of improvement in April was 10.7 fewer words per year on the 50-word list, whereas the EC group's rate of improvement was 8.6 fewer words per year. The shape of the growth curves depicted in Figure 2 indicates a pattern of increasing differences over time, and is evidenced by the higher rate of change in April for the DC group. DC children also had higher expected word-reading scores (mean intercept) in April than IC-R children ( $t = 2.26, p = .024, f = 1.03$ ), although this difference is slightly above the Bonferroni-adjusted level of alpha (i.e., .024 vs. .0167). This was a 5.1-word difference between the DC and IC-R groups in April. These differences are shown in the raw and predicted scores plotted in Figure 2.

To further examine possible group differences in word reading at the end of the school year, a two-level random-effects model was run on April word-reading scores using



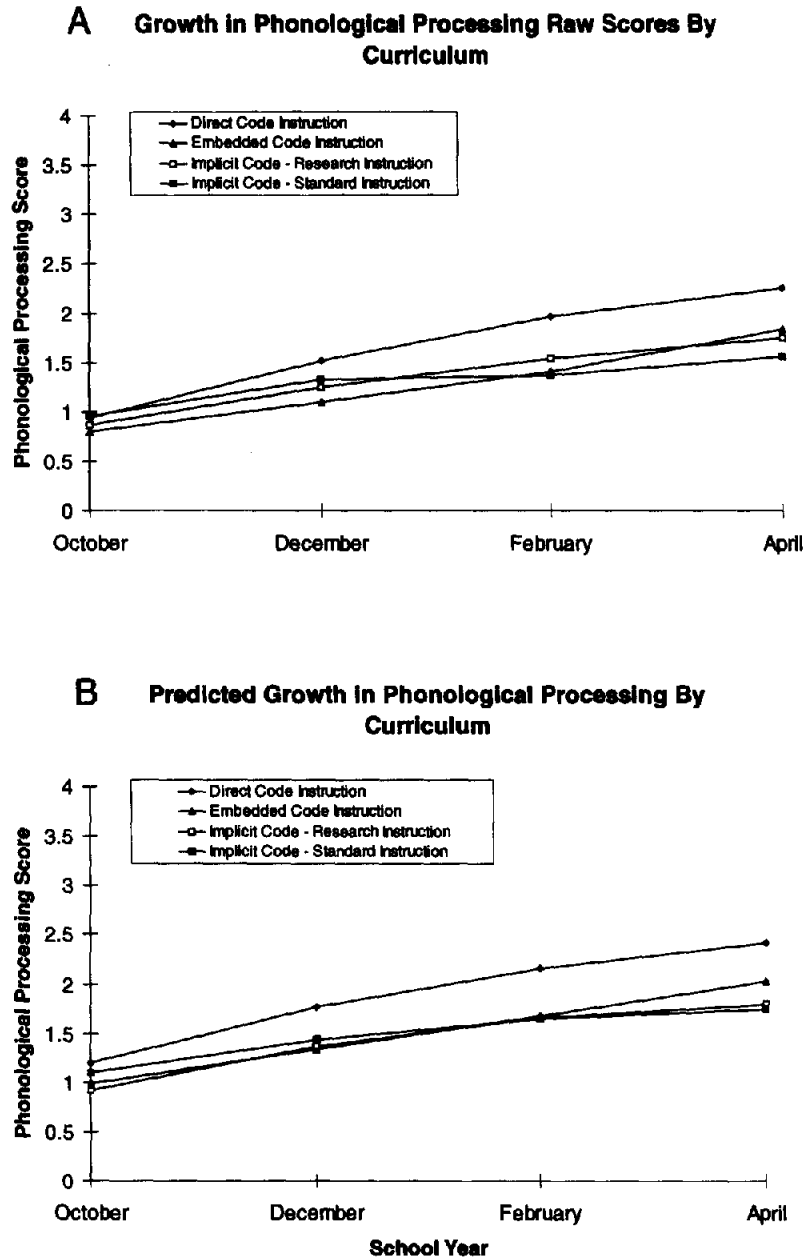


Figure 1. Growth in phonological processing raw scores by curriculum (panel a) and predicted growth in phonological processing by curriculum (panel b).

HLM-2 (Bryk & Raudenbush, 1992). We included covariates of age, verbal IQ, ethnicity, and October word-reading scores. This analysis revealed that the DC group outperformed the IC-R group,  $F(1, 165) = 10.06, p = .002, f = 1.53$ , as well as the EC group,  $F(1, 165) = 5.34, p = .022, f = 1.12$ , with no differences between the IC-R and EC groups ( $p = .37, f = 0.41$ ).

The practical significance of the slope and intercept differences is clearly apparent when examining individual cases. A relatively large percentage of children in the IC-R, IC-S, and EC curricula did not exhibit growth. As can be

seen in the frequency distributions of growth estimates in word reading shown in Figure 3, approximately 46% of the IC-R children, 44% of the EC children, and 38% of the IC-S children learned at a rate of 2.5 words or less per school year on the 50-word list compared with only 16% in the DC group. For DC children, growth in word reading does not have a large positive skew, indicating small amounts of growth characteristic of the other instructional groups.

To evaluate these patterns further, we used logistic regression to calculate the probability of a child having a predicted word-reading score in April greater than one,

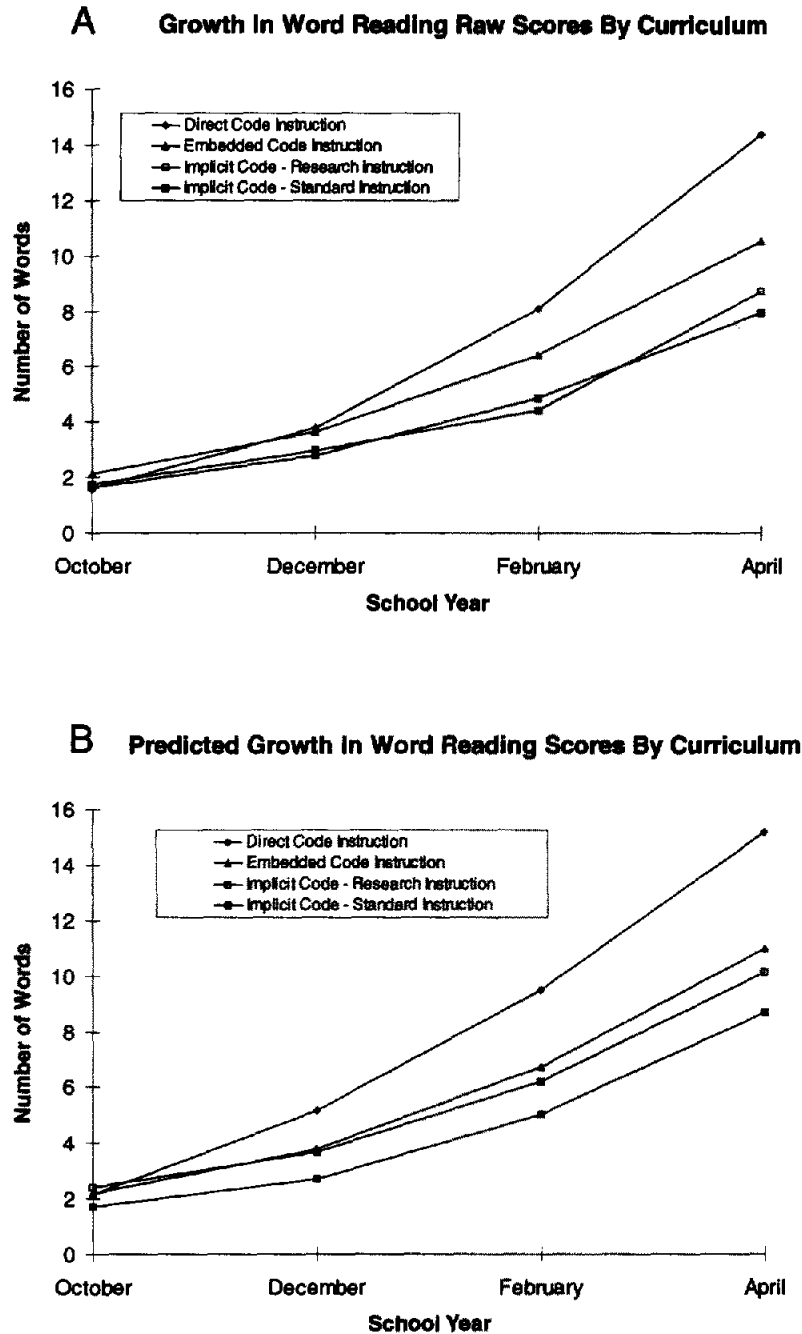


Figure 2. Growth in word reading raw scores by curriculum (panel a) and predicted growth in word-reading scores by curriculum (panel b).

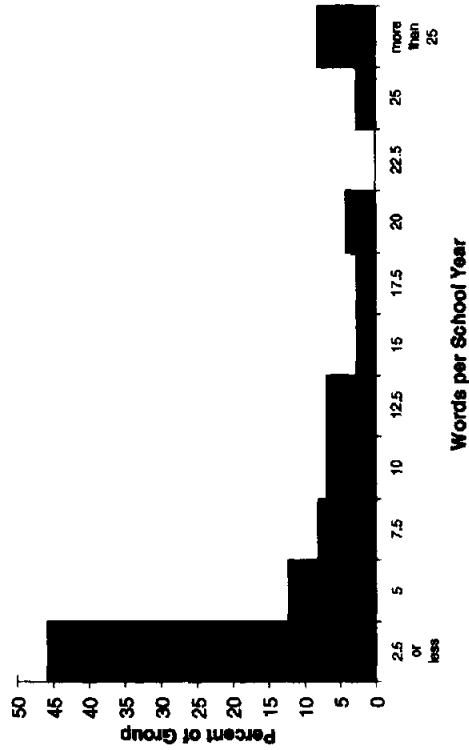
given that in October they read zero words. Included in the analysis were covariates of age and ethnicity. The results showed that DC children were 3.6 times more likely to be reading more than one word at the end of the year than IC-R children,  $\chi^2(1, N = 182) = 6.48, p = .011$  (95% confidence interval [CI] = 1.34, 9.49), and 5.2 times more likely than EC children,  $\chi^2(1, N = 182) = 10.79, p = .001$  (95% CI = 1.94, 13.80). If the criterion was two words read accurately at the end of the year, then DC children were 5.6

times more likely to be reading at that level than IC-R children,  $\chi^2(1, N = 182) = 12.74, p < .001$  (95% CI = 2.17, 14.33), and 5.2 times more likely than EC children,  $\chi^2(1, N = 182) = 11.60, p = .0007$  (95% CI = 2.014, 13.45).

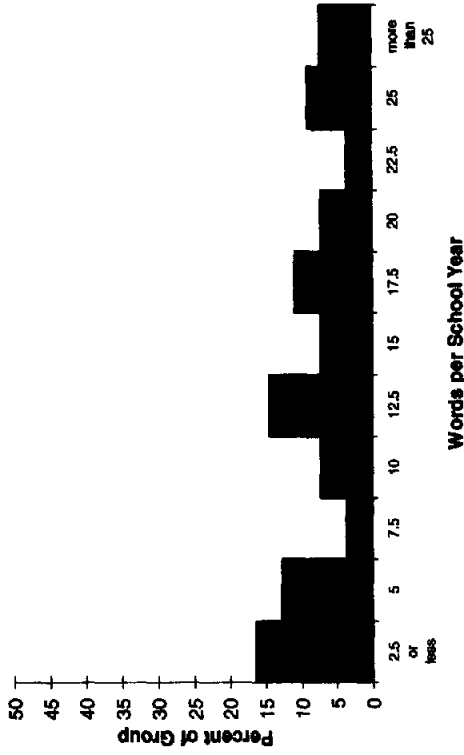
To evaluate the possible role of initial status in phonological processing in growth in word reading, October scores in phonological processing were included in a three-level analysis of word reading using HLM-3. Controlling for effects resulting from ethnicity, the phonological covariate

# Frequency Distributions of Predicted Growth in Word Reading in April

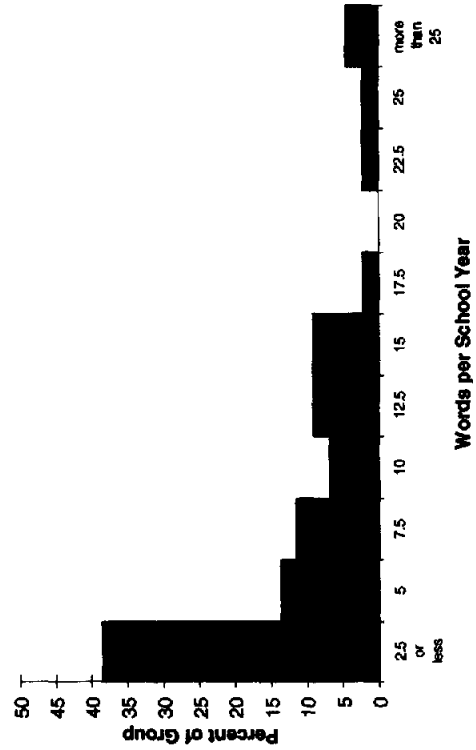
**Implicit Code - Research Instruction**



**Direct Code Instruction**



**Implicit Code - Standard Instruction**



**Embedded Code Instruction**

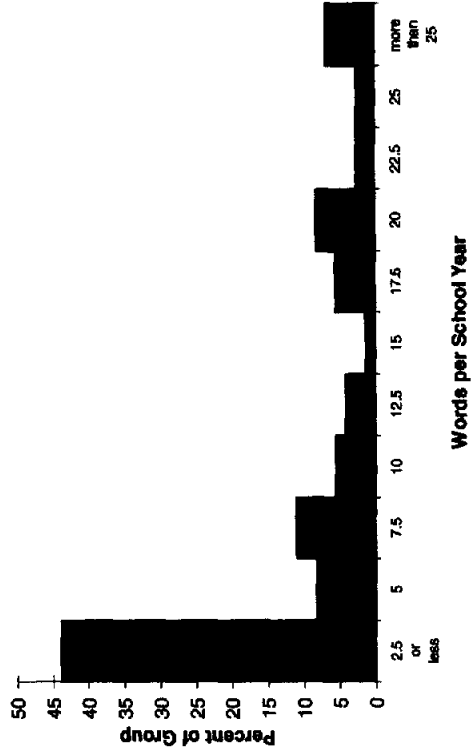


Figure 3. Frequency distributions of predicted growth in word reading.

significantly predicted rate of growth as well as April scores in word reading ( $t = 6.41$  and  $8.54$ , respectively,  $p < .001$ ).

The effects of initial phonological processing differed across instructional groups. Nevertheless, instructional group differences were similar to the model of word reading without the phonological covariate. IC-R and IC-S groups did not differ in slope or intercept ( $p > .05$ ,  $f = 0.40$  for intercepts,  $f = 0.09$  for slopes); however, there were differences among the three experimental groups. With respect to the intercept, DC children continued to have significantly higher expected scores in April than the IC-R children ( $t = 2.38$ ,  $p = .017$ ,  $f = 0.92$ ). With respect to slope, DC children continued to improve in word-reading skills at a faster rate than the IC-R children ( $t = 2.93$ ,  $p = .004$ ,  $f = 0.54$ ), whereas the difference between DC and EC, which was previously not significant at the Bonferroni-adjusted criterion, now failed to reach significance at conventional levels ( $t = 1.13$ ,  $p = .261$ ,  $f = 0.33$ ).

The differential effect of initial phonological skill on individual differences in growth of word reading is depicted in Figure 4, in which individual October scores in phonological processing are plotted separately for each group against predicted growth estimates in word reading. Generally, higher initial scores in phonological processing coincide with higher growth in word reading, and this pattern holds

for all groups. More importantly, Figure 4 shows that children who start the year with the lowest levels of phonological processing skill exhibit the lowest growth in word reading in all groups except the DC group. Indeed, some children who start the year with low phonological scores still manage to exhibit considerable growth in reading words. These children were largely in the DC instructional group, as evidenced by the vertical spread in the data points in the left side of the panel for DC and the lack of spread in the left side of the remaining three panels. The lines in the panels depict the least squares regression line relating reading growth to initial phonological processing. Although the overall test of slope differences among instructional groups was statistically significant,  $\chi^2(3, N = 252) = 7.90$ ,  $p = .048$ , none of the pairwise comparisons met the Bonferroni-adjusted critical value. Nevertheless, the generally flatter line for the DC group is precisely what one would expect if phonological processing is a determinant of growth in word reading and DC is effective in improving phonological processing. We would expect *initial* phonological processing to be less related to outcome in DC because more explicit instruction in the alphabetic code is more effective in developing phonological processing skill in all children, which thereby minimizes the importance of the level of this skill that children bring to the classroom in the fall.

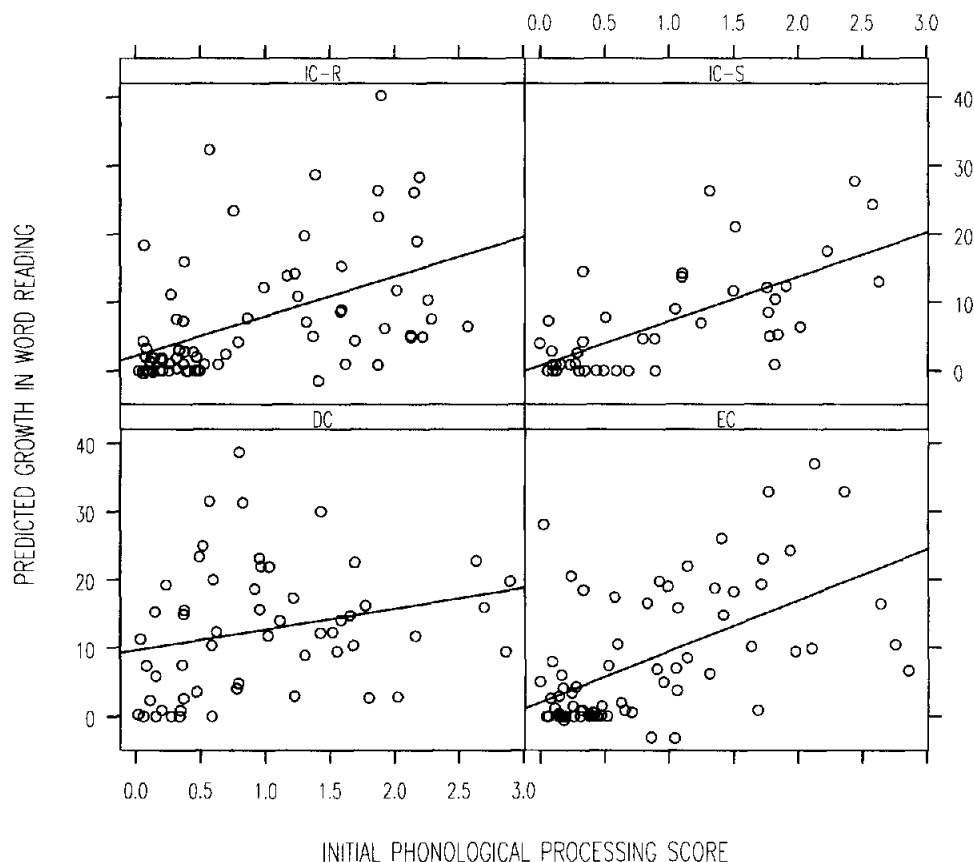


Figure 4. Plots of individual growth estimates in word reading by initial phonological processing scores and instructional group.

*Analysis of growth in vocabulary.* In the analysis of growth in vocabulary using the PPVT-R, there was no evidence for quadratic change. Rather, growth was linear ( $t = 11.22$ ,  $p < .001$ ). In addition, there were significant effects of age ( $t = 8.13$ ,  $p < .001$ ) on expected vocabulary in April, and Hispanic children had lower expected vocabulary scores in April compared with the sample average ( $t = 4.86$ ,  $p < .001$ ). Most important, there were no effects as a result of instructional group (overall effect size  $f = 0.16$ ,  $\Delta R^2 = .01$ ). Thus, IC-R, IC-S, EC, and DC children all developed to the same level and at the same rate in vocabulary (i.e., about 6.5 items on the PPVT-R per

school year), which shows that the effect of DC on cognitive skills was specific to reading and did not reflect a generic effect of intervention. This growth in vocabulary is depicted in Figure 5 in terms of raw (panel a) and predicted (panel b) scores.

### End-of-Year Achievement

Standard score means and standard deviations for the May achievement tests of reading and spelling are provided in Table 5 for each instructional group. The WJ-R Basic

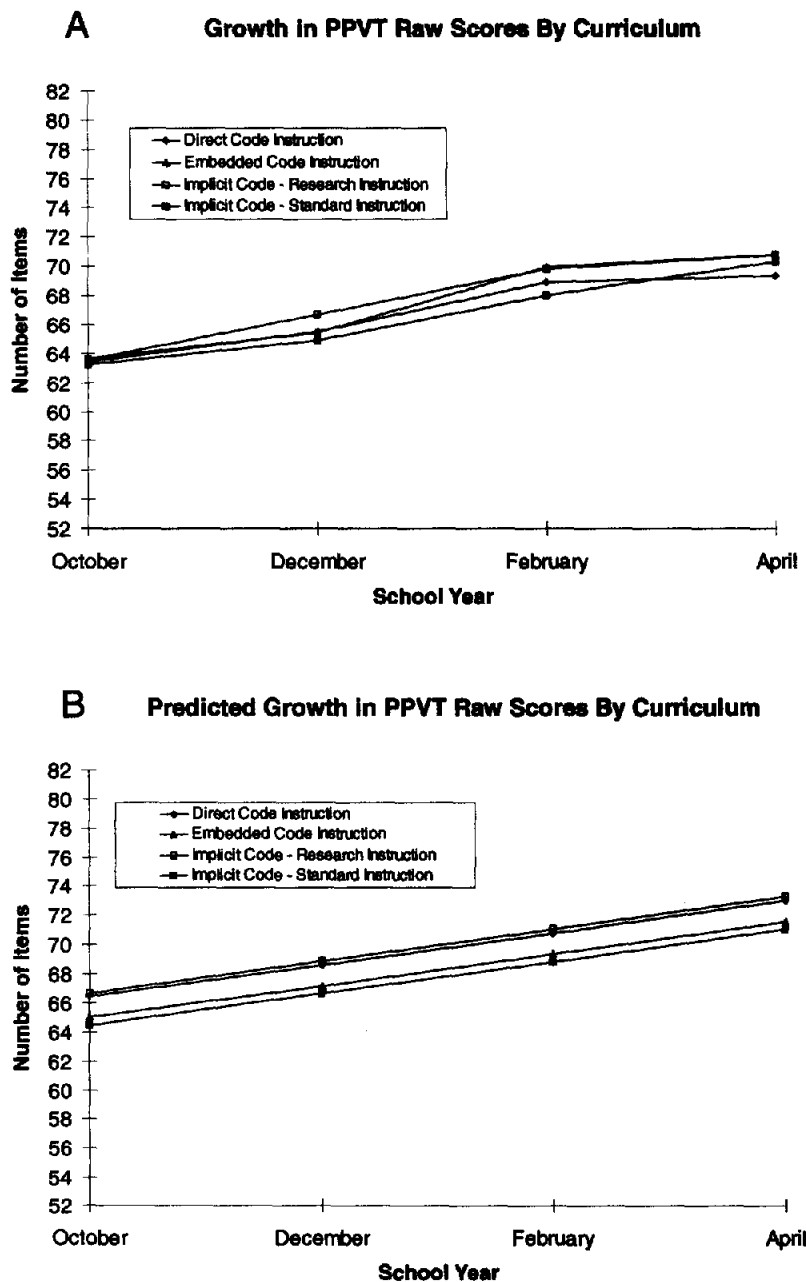


Figure 5. Growth in Peabody Picture Vocabulary Test—Revised (PPVT-R) raw scores by curriculum (panel a) and predicted growth in PPVT-R scores by curriculum (panel b).

Table 5  
Standard Score Means, Standard Deviation, and Sample Sizes on May Achievement Tests of Reading and Spelling for Four Instructional Groups

Instructional group	WJ-R Reading			FRI comprehension
	Basic	Passage comprehension	KTEA spelling	
Direct code				
<i>M</i>	96.1	96.7	85.7	81.8
<i>SD</i>	14.6	15.9	12.2	9.4
<i>n</i>	58	58	58	50
Embedded code				
<i>M</i>	88.6	91.4	82.0	80.8
<i>SD</i>	11.2	12.7	8.2	8.3
<i>n</i>	82	82	82	62
Implicit code-research				
<i>M</i>	89.6	92.0	81.6	81.5
<i>SD</i>	12.7	14.8	9.1	8.7
<i>n</i>	78	78	77	61
Implicit code-standard				
<i>M</i>	84.5	89.0	81.7	83.1
<i>SD</i>	9.7	12.1	7.6	6.9
<i>n</i>	45	45	45	34

Note. WJ-R = Woodcock-Johnson Psychoeducational Battery-Revised (Woodcock & Johnson, 1989); KTEA = Kaufman Test of Educational Achievement (Kaufman & Kaufman, 1985); FRI = Formal Reading Inventory (Wiederholt, 1986). The FRI was not administered to children who scored less than 5 points on the WJ-R Passage Comprehension.

Reading cluster is the average of the Letter-Word Identification and Word Attack (pseudoword) subtests and represents a measure of decoding. Passage Comprehension is a cloze test at the sentence level, and the FRI is a multiple-choice test based on silent narrative and expository text reading. On the basis of our previous research (Foorman et al., 1996), we did not administer the FRI to children who scored less than 5 raw score points on the WJ-R Passage Comprehension to avoid frustrating the children on the more difficult FRI.

A two-level hierarchical linear models approach using SAS PROC MIXED (SAS Institute, 1997), nesting student within teacher, was utilized to investigate instructional group differences in the May achievement scores. Significant effects of instructional group were followed up with the three post hoc contrasts of interest, using Bonferroni corrections to control the alpha level at  $p < .0167$ . Significant instructional group effects were found for the WJ-R Basic Reading cluster,  $F(3, 197) = 6.03, p = .008, f = 0.67, \Delta R^2 = .48$  and the WJ-R Passage Comprehension subtest,  $F(3, 197) = 2.75, p = .044, f = 0.40, \Delta R^2 = .64$ . Post hoc tests of the instructional effect revealed that the DC group had higher mean decoding scores than either the EC group,  $F(1, 197) = 9.41, p = .003, f = 1.17$ , or the IC-R group,  $F(1, 197) = 7.00, p = .009, f = 1.22$ , respectively. Likewise, the DC group had higher mean Passage Comprehension scores than the EC group,  $F(1, 197) = 4.76, p = .030, f = 0.72$ , but this difference was not significant at the Bonferroni-adjusted criterion. The difference between the DC and IC-R groups was not significant,  $F(1, 197) = 3.68, p = .056, f =$

0.76. Although these differences on Passage Comprehension did not meet critical alpha values, the direction of the differences is clear and the magnitude of the effects is large by typical standards. There were no instructional group differences on the KTEA Spelling or on the FRI ( $p > .05$ , overall  $f_s = 0.38$  and  $0.20, \Delta R^2 = .22$  and undefined, respectively). The FRI was too difficult for these children, as is apparent from the low means of Table 6 and the fact that a sizable number of children in each group (i.e., 14% of DC and about 24% of the other groups) were not administered the FRI because they did not meet the criterion of scoring at least 5 raw score points on the WJ-R Passage Comprehension.

We used logistic regression to calculate the probability of a child having a May WJ-R decoding score below the 25th percentile, a usual diagnostic criteria for a reading disability (Fletcher et al., 1994). IC-S and IC-R children did not differ from each other. However, IC-R children were 2.4 times as likely as DC children to score below the 25th percentile,  $\chi^2(1, N = 262) = 5.21, p = .02$  (95% CI = 1.3, 4.1), and EC children were 3.1 times as likely as DC children to score below the 25th percentile,  $\chi^2(1, N = 262) = 10.09, p = .002$  (95% CI = 1.5, 6.4).

#### *Analyses of Attendance, Perceived Self-Competence, Attitudes, Behavior, and Environmental Variables*

Instructional groups did not differ in school attendance, in perceived self-competence on the Harter scales, or in teacher identification of emotional, behavioral, or family problems on the end-of-year evaluation. However, instructional groups significantly differed in reading attitudes (but not experience),  $F(3, 257) = 4.29, p = .006$ . The IC-R group had more positive attitudes toward reading than the DC group,  $F(1, 257) = 6.29, p = .013$ , and the IC-S group,  $F(1, 257) = 11.12, p = .001$ . Questions related to the degree to which the child likes people to read to him or her, likes or does not like to read books by him or herself, thinks learning to read is hard or easy, likes or does not like school, likes or does not like to watch television, and has a parent, grandparent, guardian, or sibling who likes or does not like to read.

Means and standard deviations are provided in Table 6 for the six scales of the MIT. ANOVAs on the six scales revealed instructional group differences on all scales but the Attention scale. Using Bonferroni adjustment for alpha (.05/6 scales = .0083), pairwise post hoc contrasts revealed the IC-S group to be significantly different from the other groups. With respect to the activity scale, the IC-S group had significantly higher activity ratings (e.g., out of chair, restless, distractible) than the IC-R group,  $F(1, 271) = 8.81, p = .003$ , and the DC group,  $F(1, 271) = 7.95, p = .005$ . The IC-S group had significantly poorer Adaptability scores (e.g., gets upset and cannot tolerate changes, transition problems, long time to settle down) compared with the IC-R group,  $F(1, 271) = 14.05, p = .0002$ , and the EC group,  $F(1, 271) = 8.66, p = .004$ . The IC-S group also had significantly poorer Social scores (e.g., calls out in class, easily frustrated) relative to the IC-R group,  $F(1, 271) = 11.08, p < .001$ . On the Academic scale, the IC-S group had significantly lower academic ratings relative to the EC group,  $F(1,$

Table 6  
Means, Standard Deviations, and *p* Values for the Six Scales of the Multi-Grade Inventory for Teachers for Four Instructional Groups

Scales	Direct code			Embedded code			Implicit code-research			Implicit code-standard			<i>p</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	
Academic	3.26	0.41	60	3.11	0.41	86	3.26	0.42	85	3.39	0.33	47	.002
Activity	2.96	1.51	60	3.14	1.48	86	2.97	1.47	85	3.77	1.42	47	.020
Adaptability	2.89	0.77	60	2.82	0.93	86	2.70	0.80	85	3.28	0.89	47	.003
Attention	3.58	0.85	60	3.59	0.79	86	3.38	0.84	85	3.65	0.77	47	.189
Language	2.86	0.78	60	2.85	0.72	86	2.68	0.67	85	3.09	0.56	47	.020
Social	3.27	0.47	60	3.25	0.54	86	3.14	0.59	85	3.47	0.56	47	.010

271) = 14.49,  $p = .0002$ . With respect to the Language scale, the IC-S group had significantly more problems (e.g., trouble expressing thought, difficult to understand) compared with the IC-R group,  $F(1, 271) = 10.43$ ,  $p < .001$ , respectively. Thus, the IC-S teachers perceived that their students had significantly more behavioral and academic problems compared with the IC-R, DC, and EC teachers.

### Discussion

The results of this research clearly indicate that early instructional intervention makes a difference for the development and outcomes of reading skills in first- and second-grade children at risk for reading failure. However, the results also demonstrate that not all instructional approaches have the same impact. Children who were directly instructed in the alphabetic principle improved in word-reading skill at a significantly faster rate than children indirectly instructed in the alphabetic principle through exposure to literature. Furthermore, 46% of the children in the IC research group and 44% of the EC group exhibited no demonstrable growth in word reading compared with only 16% in the DC group.

These performance differences were due to instruction, not to behavioral or affective differences among these groups. The only differences on the behavioral measures involved the IC-S condition, not a surprising finding given that the vast majority of IC-S children came from one school described as "tough." Because this school was the "unseen control," we did not monitor classroom reading instruction and, therefore, cannot determine the extent to which these perceived behavioral and academic problems may have been a consequence of poor classroom instruction. There were no behavioral differences among the three research conditions. Similarly, although outcomes varied across classrooms, measured characteristics of the teachers did not relate significantly to outcome. Generally, teachers' attitude toward and compliance with instructional practices were very good across instructional groups, and the amount of time devoted to reading and language arts instruction was comparable.

Children in all instructional groups with higher initial status in phonological processing skills in October exhibited growth in word-reading skills. However, children in the DC group who had low initial status in phonological processing skills also appeared to show more growth in word-reading

skills than children with low phonological processing scores in the other instructional groups. Hence, the fact that the DC approach used in this study included explicit instruction in phonemic awareness appeared to facilitate word-reading development for children who started the year with low scores in this crucial precursor skill to reading. This shows not only that problems with phonological processing are related to poor reading skills in these culturally and linguistically diverse children, but that greater changes in phonological processing skills and word-reading ability occurred when these children were provided a curriculum that included explicit instruction in the alphabetic principle. The finding that phonological processing moderated growth in word reading suggests that the changes were due to the nature of the instruction and not to the greater scripting of the DC approach. Nevertheless, future studies should compare the DC program used in this study with other DC programs that vary in the degree of scriptedness to evaluate this possibility. Also, the onset-rime component of the EC intervention was scripted in the sense that spelling patterns were systematically presented. Hence, it is not surprising that performance of the EC group tended to fall between that of the DC and IC groups.

Instructional group differences in end-of-year achievement after the first year were clearly apparent: The direct instruction group approached national average on decoding (43rd percentile) and passage comprehension (45th percentile) compared with the IC-R group's means of 29th percentile and 35th percentile, respectively. (EC group means were 27th percentile and 33rd percentile, respectively). Although the differences in decoding skills were robust, mean differences on the Passage Comprehension test did not meet the critical value of alpha adopted for this study. However, our approach was designed to minimize Type I errors and was conservative. The mean differences on this measure of reading comprehension were large; effect sizes were also large, favoring the DC group. Furthermore, logistic regression revealed that children in the IC-R and EC groups were much more likely to score below the 25th percentile on the standardized decoding test than children in the DC instruction group. Scores below the 25th percentile are often used to indicate reading disability on the basis of traditional diagnostic criteria (Fletcher et al., 1994).

In this study, there were no effects of student-teacher ratio

or nature of content in the tutoring component. However, the student-teacher ratio was not a constant 1:1 or small group variable because of teachers' need to reconstitute groups to adjust for behavioral or learning differences. Therefore, we do not see our results as inconsistent with the research supporting the benefits of one-on-one tutoring (e.g., Wasik & Slavin, 1993). Future research should continue to study the benefits of having tutorial content match or not match classroom instruction. Having the content of tutorial match the curriculum facilitates communication between classroom teacher and tutor and ensures continuity of treatment for the child (see Slavin et al., 1996). However, many tutoring programs are springing up around the United States in response to the America Reads challenge, and these programs entail training that is divorced from classroom instruction. Disconnected instructional programs have been shown to be ineffective for high-risk students (Allington, 1991).

As with any other intervention study, longer term follow-up with these children is clearly indicated to assess whether the gains in decoding skills continue to accelerate in DC instruction and whether there are longer term effects in other aspects of the reading process. For example, in spite of differences in decoding skills, the IC group had more positive attitudes toward reading, a finding consistent with other research (e.g., Stahl, McKenna, & Pagnucco, 1994). It is possible that these positive attitudes toward reading, although not associated with higher reading performance in beginning reading, may sustain motivation to improve reading skill as the student matures. Another interesting question is whether the sequence of instructional method makes a difference in growth and outcomes in reading. For example, do children who receive explicit instruction in the alphabetic principle in Grade 1 and subsequent implicit instruction show greater gains than children who continue in explicit instruction? Similarly, can direct instruction in alphabetic and orthographic rules in Grade 2 ameliorate the lack of growth in reading experienced in Grade 1 by children who received either of the IC approaches? The effects of EC instruction may require a longer period of time for benefits to be realized. The large individual differences in the EC group support findings from previous research (e.g., Ehri & Robbins, 1992; Foorman, 1995a) that some decoding skill is needed before known orthographic rimes are spontaneously used to read unknown words by analogy. At the same time, it may take more time for children to use the spelling patterns taught in the EC program. Hence, DC instruction may be more efficient and lead to more rapid initial rates of growth, but it is possible that the effects of an EC approach are cumulative so that longer term outcomes are not different. The critical issue is the extent to which the earlier development of decoding skills achieved with explicit instruction is associated with improvement in reading comprehension and spelling, which remains an open issue.

The positive effects of DC instruction did not generalize to all academic areas. Instructional groups did not differ in spelling achievement, and the average spelling scores were not impressive. The measure of text reading had a floor effect. Subsequent assessments will, it is hoped, show

greater transfer of word-reading skills to the text reading in measures such as the FRI as well as measures such as the WJ-R Passage Comprehension. Finally, 90 children who were eligible for Title 1 and who received the classroom intervention were not included in these analyses because they did not receive tutorial services. These children were better readers than the children in these analyses at baseline. Analyses that included these 90 children did not alter the pattern of results.

It is also important to keep in mind that the classroom curricula used in this study took place in a print-rich environment with a significant literature base. Instructional programs that provided only phonological awareness or phonics lessons were not used because it was not likely that such training would generalize to actual reading and spelling skills. In the DC condition, as in other intervention studies with demonstrable efficacy with poor readers (Torgesen, 1997; Vellutino et al., 1996), explicit instruction in the alphabetic principle was separated from the literature component, but both components were provided. The opportunity to apply what is learned in this component is most likely critical for ensuring that the instruction generalizes.

The results of this study underscore the value of research informed by contemporary hypotheses regarding the interconnection between language and reading. Previous research has demonstrated the effectiveness of direct instruction in the alphabetic principle with beginning readers from middle-class schools (Foorman, 1995a, 1995b; Foorman, Francis, Novy, & Liberman, 1991; Vellutino et al., 1996) as well as with disabled readers (Torgesen, 1997; Vellutino et al., 1996). Although the effects of tutorial interventions in this study were overshadowed by the strong effects of classroom instruction, other research with severely disabled readers indicates the merits of intensive one-to-one intervention with students (Torgesen, 1997; Vellutino et al., 1996). Future studies should also evaluate entire classrooms, not just Title 1 children, and compare the DC program in this study with other curricula providing DC. Depending on the results, it may well be possible to prevent reading failure for large numbers of children if beginning instruction explicitly teaches the alphabetic principle.

## References

- Adams, M. J. (1990). *Beginning to read*. Cambridge, MA: MIT Press.
- Agronin, M. E., Holahan, J. M., Shaywitz, B. A., & Shaywitz, S. E. (1992). The Multi-Grade Inventory for Teachers. In S. E. Shaywitz & B. A. Shaywitz (Eds.), *Attention deficit disorder comes of age* (pp. 29-67). Austin, TX: PRO-ED.
- Allington, R. L. (1991). Children who find learning to read difficult: School responses to diversity. In E. H. Hiebert (Ed.), *Literacy for a diverse society* (pp. 237-252). New York: Teachers College Press.
- Archambault, R. X. (1989). Instructional setting and other design features of compensatory education programs. In R. E. Slavin, N. L. Karweit, & N. A. Madden (Eds.), *Effective programs for students at risk* (pp. 220-263). Needham Heights, MA: Allyn & Bacon.



- Aukerman, R. C. (1984). *Approaches to reading*. New York: Wiley.
- Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly*, 26, 49–66.
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature*, 301, 419–421.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101, 147–158.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.
- Bryne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology*, 83, 451–455.
- Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology*, 85, 104–111.
- 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.
- Carroll, J., Davies, P., & Richman, B. (1971). *The American heritage word frequency book*. Boston, MA: Houghton Mifflin.
- Chall, J. S. (1983). *Learning to read: The great debate* (2nd ed.). New York: McGraw-Hill.
- Clay, M. (1991). *Reading recovery: A guidebook for teachers in training*. Portsmouth, NH: Heinemann.
- Dahl, K. L., & Freppon, P. A. (1995). A comparison of innercity children's interpretations of reading and writing instruction in the early grades in skills-based and whole language classrooms. *Reading Research Quarterly*, 30, 50–75.
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.
- Ehri, L. C., & Robbins, C. (1992). Beginners need some decoding skill to read by analogy. *Reading Research Quarterly*, 27, 13–26.
- Englemann, S., & Bruner, E. C. (1995). *SRA Reading Mastery Rainbow Edition*. Chicago, IL: SRA/McGraw-Hill.
- Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., Francis, D. J., Fowler, A. E., & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology*, 86, 6–23.
- Foorman, B. R. (1994). The relevance of a connectionist model of reading for "the great debate." *Educational Psychology Review*, 6, 25–47.
- Foorman, B. R. (1995a). Practiced connections of orthographic and phonological processing. In V. W. Berninger (Ed.), *The varieties of orthographic knowledge: II. Relationships to phonology, reading, and writing* (pp. 377–419). Dordrecht, the Netherlands: Kluwer Academic.
- Foorman, B. R. (1995b). Research on "the great debate": Code-oriented versus whole language approaches to reading instruction. *School Psychology Review*, 24, 376–392.
- Foorman, B. R., Francis, D. J., Beeler, T., Winikates, D., & Fletcher, J. M. (1997). Early interventions for children with reading problems: Study designs and preliminary findings. *Learning Disabilities: A Multi-Disciplinary Journal*, 8, 63–71.
- Foorman, B. R., Francis, D. J., Fletcher, J. M., & Lynn, A. (1996). Relation of phonological and orthographic processing to early reading: Comparing two approaches to regression-based, reading-level-match design. *Journal of Educational Psychology*, 88, 639–652.
- 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.
- Foorman, B. R., Francis, D. J., Shaywitz, S. E., Shaywitz, B. A., & Fletcher, J. M. (1997a). The case for early reading interventions. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 243–264). Hillsdale, NJ: Erlbaum.
- Foorman, B. R., Francis, D. J., Winikates, D., Mehta, P., Schatschneider, C., & Fletcher, J. M. (1997b). Early intervention for children with reading disabilities. *Scientific Studies of Reading*, 1, 255–276.
- Francis, D. J., Fletcher, J. M., Stuebing, K. K., Davidson, K. C., & Thompson, N. M. (1991). Analysis of change: Modeling individual growth. *Journal of Consulting and Clinical Psychology*, 59, 27–37.
- Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, 88, 3–17.
- Gough, P. B., & Hillinger, M. L. (1980). Learning to read: An unnatural act. *Bulletin of the Orton Society*, 30, 179–196.
- Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development*, 53, 87–97.
- Harter, S., & Pike, R. (1984). The Pictorial Scale of Perceived Competence and Social Acceptance for young children. *Child Development*, 55, 1969–1982.
- Hiebert, E. H., Colt, J. M., Catto, S. L., & Gary, E. C. (1992). Reading and writing of first-grade students in a restructured chapter 1 program. *American Educational Research Journal*, 29, 545–572.
- Iversen, S., & Tunmer, W. (1993). Phonological processing skills and the reading recovery program. *Journal of Educational Psychology*, 85, 112–126.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80, 437–447.
- Kaufman, A. S., & Kaufman, N. L. (1985). *Kaufman Test of Educational Achievement*. Circle Pines, MN: American Guidance Service.
- Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1–33). Ann Arbor, MI: University of Michigan Press.
- Lyon, G. R. (1995). Towards a definition of dyslexia. *Annals of Dyslexia*, 45, 3–27.
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparisons perspective*. Belmont, CA: Wadsworth.
- Olson, R. K., Wise, B., Ring, J., & Johnson, M. (1997). Computer-based remedial training in phoneme awareness and phonological decoding: Effects on the post-training development of word recognition. *Scientific Studies of Reading*, 1, 235–254.
- Open Court Reading. (1995). *Collections for young scholars*. Chicago and Peru, IL: SRA/McGraw-Hill.
- Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Selzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly*, 29, 8–38.
- Puma, M. J., Jones, C. C., Rock, D., & Fernandez, R. (1993). *Prospects: The congressionally mandated study of educational growth and opportunity (interim report)*. Bethesda, MD: ABT Associates.

- Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, 90, 726-748.
- Routman, R. (1991). *Invitations: Changing as teachers and learners: K-12*. Portsmouth, NH: Heinemann.
- SAS Institute. (1997). *SAS/STAT software: Changes and enhancements through release 6.12*. Cary, NC: Author.
- Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly*, 30, 958-996.
- Shaywitz, S. E., Fletcher, J. M., & Shaywitz, B. A. (1994). Issues in the definition and classification of attention deficit disorder. *Topics in Language Disabilities*, 14, 1-25.
- Slavin, R. E., Karweit, N. L., & Madden, N. A. (1989). *Effective programs for students at risk*. Boston: Allyn & Bacon.
- Slavin, R. E., Karweit, N. L., & Wasik, B. A. (1994). *Preventing early school failure: Research on effective strategies*. Boston: Allyn & Bacon.
- Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1996). *Every child, every school: Success for all*. Thousand Oaks, CA: Corwin Press.
- Stahl, S. A., McKenna, M. C., & Pagnucco, J. R. (1994). The effects of whole-language instruction: An update and a reappraisal. *Educational Psychologist*, 29, 175-186.
- Stanovich, K. E. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. *Developmental Review*, 10, 72-100.
- Stanovich, K. E., & Siegel, L. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology*, 86, 24-53.
- Torgesen, J. K. (1997). The prevention and remediation of reading disabilities: Evaluating what we know from research. *Journal of Academic Language Therapy*, 1, 11-47.
- Vellutino, F. R. (1991). Introduction to three studies on reading acquisition: Convergent findings on theoretical foundations of code-oriented versus whole-language approaches to reading instruction. *Journal of Educational Psychology*, 83, 437-443.
- Vellutino, F. R., Scanlon, D. M., Sipay, E., Small, S., Pratt, A., Chen, R., & Denckla, M. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, 88, 601-638.
- Wagner, R., 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.
- Wasik, B., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, 28, 178-200.
- Weaver, C. (1994). *Understanding whole language: From principles to practice*. Portsmouth, NH: Heinemann.
- Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: Psychological Corporation.
- Wiederholt, L. J. (1986). *Formal Reading Inventory*. Austin, TX: PRO-ED.
- Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psychoeducational Battery-Revised*. Allen, TX: DLM Teaching Resources.

## Appendix A

### Spelling Patterns for Embedded Code Instruction (Sequenced From Left to Right)

Sequence for Grade 1					Sequence for Grade 2				
_at	_ad	_an	_am	_ap	_ath	_uff	_oss	_ess	_all
_on	_op	_ot	_og	_id	_ill	_ell	_ax	_act	_alt
_ig	_it	_in	_up	_un	_and	_amp	_ant	_ast	_anch
_ug	_ut	_ud	_et	_ed	_ang	_ank	_ask	_atch	_ox
_en	_ep	_go, no	_ain	_aint	_oft	_ond	_ong	_onk	_ook
_ait	_ay	_ame	_ake	_ade	_oot	_oom	_ood	_oon	_ix
_ate	_ale	_ave	_ane	_ace	_ift	_ink	_int	_ilt	_ilk
_me	_my	_ee	_eed	_earn	_ist	_isp	_ing	_irst	_itch
_ean	_ease	_eet	_ea	_ead	_ump	_ust	_unt	_ulp	_unch
_eal	_eel	_een	_eep	_eat	_ung	_unk	_ept	_end	_elt
_each	_ail	_ite	_ile	_ine	_ench	_est	_elp	_eft	_elf
_ive	_ice	_ime	_ike	_ide	_ent	_oil	_eive	_ief	_ar
_ome	_oke	_ose	_ope	_igh	_arm	_arn	_orm	_orn	_ern
_ight	_ind	_old	_ue	_oa	_ir	_ird	_urn		
_oad	_oat	_ie	_ow	_own					

## Appendix B

## Instructional Components Used as Criteria for Compliance

Direct code components	
1. Phonemic awareness	5. Writing
2. Use of anthology	6. Spelling dictation
3. Phonics, phonics review	7. Workshop
4. Guided and independent exploration	8. Use of workbook materials
Embedded code components	
1. Make-and-break activities	5. Writing (shared, independent)
2. Reading (shared, choral-echo, guided, readers' circle, independent)	6. Morning message, daily edit
3. Strategy instruction	7. Running record
4. Frame target word, extend pattern, review phonemic awareness	8. Home reading
Implicit code components	
1. Shared reading	5. Writing workshop, process
2. Guided reading	6. Integrated curriculum
3. Responses to and extensions of literature	7. Print-rich environment
4. Phonics instruction in context	8. Spelling instruction, workshop based on strategies and meaningful context

## Appendix C

## Teacher Attitude Survey

1. If you were responsible for curriculum decisions in your district, would you recommend that resources (materials, staff development, etc.) be provided for this intervention in the future?	5. How close is the match between the intervention you are delivering and your own beliefs about how to teach children to read?
2. Would you recommend the intervention you are using to a colleague?	a. An exact match. This is the way I already teach.
3. Would you recommend the intervention for use with all age-appropriate children?	b. Very similar. I agree with most aspects of the intervention.
4. Would you recommend the intervention for children with special needs?	c. Somewhat similar. I agree with some aspects of the intervention.
	d. Not similar at all. My beliefs about the teaching of reading are contradictory to those of the intervention.

*Note.* Responses to the first four questions were based on a scale ranging from 1 (*definitely yes*) to 5 (*definitely no*).

Received February 6, 1997  
Revision received July 1, 1997  
Accepted September 16, 1997 ■