

Is Reading Important in Reading-Readiness Programs? A Randomized Field Trial With Teachers as Program Implementers

Douglas Fuchs, Lynn S. Fuchs, Anneke Thompson,
Stephanie Al Otaiba, Loulee Yen,
Nancy J. Yang, and Mary Braun
Peabody College of Vanderbilt University

Rollanda E. O'Connor
University of Pittsburgh

The purpose of the present study was to examine the effectiveness and feasibility of phonological awareness training, with and without a beginning decoding component. Thirty-three teachers in 8 urban schools were assigned randomly within their schools to 3 groups: control, phonological awareness training, and phonological awareness training with beginning decoding instruction and practice. Following training, teachers in the 2 treatment groups conducted the treatments for about 20 weeks. In each teacher's class, pre- and posttreatment data were collected on 12–14 children ($N = 404$); 312 children were tested again the following fall. At the end of kindergarten, the 2 treatment groups performed comparably and outperformed controls on the phonological awareness measures. On alphabetic (reading and spelling) tasks, however, the group participating in phonological awareness training with beginning decoding instruction did better than the other 2 groups. In the fall of the next year, many of these between-group differences remained but were less impressive. Implications are discussed for bridging research and practice.

Phonological awareness is “very hot.” So say 25 experts from the United States, Australia, and Canada who were asked to rate 27 reading-related topics for *Reading Today*, a publication of the International Reading Association (Cassidy & Wenrich, 1998). In support of the experts' rating, one should consider the following:

1. The National Research Council's Committee on the Prevention of Reading Difficulties in Young Children recommended that “activities that direct young children's attention to the phonological structure of spoken words . . . and that highlight the relations between print and speech” be included in early childhood programs (Snow, Burns, & Griffin, 1998, p. 321).

2. The Houston public schools recently spent millions of dollars to provide all K–3 teachers with materials and 30 hr of training so

they can conduct phonemic awareness lessons and other activities thought to boost reading development (see Bradley, 1998).

3. The Reading Excellence Act, a \$260 million program targeting 500,000 children in Grades K–3, says that reading covers six vital areas, one of which is phonemic awareness (International Reading Association, 1999).

4. In 1998, the National Institute for Child Health and Human Development (NICHD) and the Office of Special Education Programs in the U.S. Department of Education cosponsored the National Summit on Research in Learning Disabilities to help disseminate research showing causal connections between phonological awareness and beginning reading (Lyon, n.d.).

5. The seven-page cover story in a November 22, 1999, issue of *Newsweek* was on dyslexia. It highlighted NICHD researchers' work on phonological awareness and concluded by recommending in part that parents wishing to help their children learn to read should “lobby for phonemic-awareness testing and . . . instruction” (Kantowitz & Underwood, 1999, p. 78).

Douglas Fuchs, Lynn S. Fuchs, Anneke Thompson, Stephanie Al Otaiba, Loulee Yen, Nancy J. Yang, and Mary Braun, Department of Special Education, Peabody College of Vanderbilt University; Rollanda E. O'Connor, Department of Instruction and Learning, University of Pittsburgh.

This research was supported in part by Grants H324V980001 and H023C0251 from the Office of Special Education Programs in the U.S. Department of Education and by Core Grant HD 15052 from the National Institute of Child Health and Development. The article does not necessarily reflect the positions or policies of these funding agencies, and no official endorsement by them should be inferred. Portions of this article were presented at the annual meeting of the American Educational Research Association in Montreal, Canada, the Biennial Conference on Evidence-Based Policies and Indicator Systems in Durham, England, and the Pacific Coast Research Conference in La Jolla, California (all in 1999).

We thank Steve Graham, Karen Harris, Joe Jenkins, Ingrid Oxaal, and Joanna Williams for their help in the planning stage of the study. We also thank our study teachers from whom we learned much about kindergarten.

Correspondence concerning this article should be addressed to Douglas Fuchs, Box 328 Peabody, Vanderbilt University, Nashville, Tennessee 37203. Electronic mail may be sent to doug.fuchs@vanderbilt.edu.

Explaining “Very Hot”

Why has phonological awareness grabbed our attention so? We suspect there are many explanations; we offer four, moving from the general to the specific. First, everyone agrees reading is pivotal to virtually all school learning. Second, many children are not learning to read. On the National Assessment of Educational Progress, 40% of fourth graders read below an established standard (see Snow et al., 1998). Approximately 4.5 million African American students and 3.3 million Hispanic children “were reading very poorly in fourth grade” (Snow et al., 1998, p. 97).

Third, phonological awareness is often equated with early intervention. It is widely acknowledged that reading failure begins early. When young children fail to learn to read, they begin to dislike reading, they read less than their more skilled classmates,

and as a consequence, they lose an important means of gaining vocabulary, background knowledge, and information about text structure (e.g., Juel, 1996). The (word-) poor get poorer as the (word-) rich get richer (see Stanovich, 1986). Juel (1988) estimated that the probability of a poor reader at the end of Grade 1 remaining a poor reader at the end of Grade 4 is .88. Hence, reading failure can set in quickly and thereafter can become difficult to remediate. This would appear to argue for intervening before children experience sustained failure and lose their self-confidence and enthusiasm for learning (e.g., Juel, 1996). Many American policymakers and researchers are bullish on early intervention programs like Head Start because they believe they work and, in the long run, save taxpayers money (e.g., Snow et al., 1998; The Consortium on Renewing Education, 1998).

Fourth, an impressive amount of correlational and experimental evidence links phonological awareness to reading. Typical (non-disabled) kindergartners with relatively strong phonological awareness read better in subsequent grades than do classmates who show comparatively weak phonological awareness in kindergarten (e.g., Bradley & Bryant, 1983; Bryant, MacLean, Bradley, & Crossland, 1990; Juel, Griffith, & Gough, 1986; Lundberg, Olofsson, & Wall, 1980; Stanovich, Cunningham, & Cramer, 1984), and the strength of this correlation endures after controlling for intelligence, vocabulary, letter knowledge, memory, or social class (e.g., Bradley & Bryant, 1985; Goswami & Bryant, 1990; Share, Jorm, MacLean, & Matthews, 1984; Torgesen, Wagner, & Rashotte, 1994; Wagner & Torgesen, 1997; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993). In other correlational studies, children with severe reading problems have demonstrated poor phonological awareness skills (e.g., Bradley & Bryant, 1983; Vellutino & Scanlon, 1987).

Experimental work on phonological awareness is typically referred to as the "training studies." At least 60 of these studies are described in the literature; they usually involve preschool and kindergarten children, and they can be divided into five types: Investigators either (a) train only phonological awareness and estimate program effectiveness exclusively on phonological awareness measures (e.g., Content, Kolinsky, Morais, & Bertelson, 1986; Olofsson & Lundberg, 1983; Rosner, 1974); (b) train only phonological awareness and use both phonological awareness and beginning reading measures to gauge treatment effects (e.g., Brennan & Ireson, 1997; Lundberg, Frost, & Peterson, 1988; Olofsson & Lundberg, 1985; Schneider, Kuspert, Roth, & Vise, 1997; Torneus, 1984; Weiner, 1994); (c) use substitutes for, or analogues of, letters in letter-sound correspondence instruction together with phonological awareness training (e.g., Cunningham, 1990; Fox & Routh, 1984; Vellutino & Scanlon, 1987); (d) use real letters for letter-sound correspondence training in combination with phonological awareness training (e.g., Ball & Blachman, 1991; Bradley & Bryant, 1983); or (e) integrate phonological awareness training with beginning reading instruction (e.g., Hatcher, Hulme, & Ellis, 1994; Wallach & Wallach, 1976; Williams, 1980).

Overall, these researchers' experimental work indicates that phonological awareness can be trained; the training can produce a positive, albeit small, effect on reading development; and its influence can be enhanced when integrated with letter-sound or beginning reading instruction. Such findings have led many scholars (e.g., Adams, 1990; Juel, 1988; Snow et al., 1998) and policymakers (e.g., Lyon, n.d.) to recommend inclusion of phonolog-

ical awareness elements in reading-readiness instruction before first grade.

Generalizability of the Training Studies

An important and obvious assumption is that the correlational and experimental training studies represent a requisite database—a critical mass of knowledge—to call for important modifications in preschool and kindergarten (and first-grade) reading-readiness programs. Such changes would entail the redesigning of preservice and inservice professional development and the redeployment of technical assistance money from governmental agencies. Despite a clear need for action and a desire no doubt felt by many researchers to provide help—not incidentally proving the real-world value of research to those who think it has little value (see Kaestle, 1993)—we must ask whether the training studies have produced reading-readiness programs that are indeed ready for export to classrooms. This concern was prompted by an impression, expressed by others before us (e.g., Blachman, Ball, Black, & Tangel, 1994; Byrne & Fielding-Barnsley, 1995; O'Connor, Notari-Syverson, & Vadasy, 1996; Troia, 1999), that most training studies have been conducted outside classrooms and with individual children or small groups of students or have been conducted inside classrooms but with research staff, rather than teachers, conducting the training.

How many training studies involved teachers as program implementers? To answer this question, we read qualitative (Troia, 1999) and quantitative (Bus & van IJzendoorn, 1999) reviews of the training studies and all articles referenced in the reviews that (a) explored the effectiveness of phonological awareness training, decoding training, or both; (b) involved preschool or kindergarten children; and (c) were published in peer-review journals. We then hand searched seven scholarly journals published for the years 1985 to 1999, inclusive, looking for similar studies. We searched *Child Development*, *Exceptional Children*, *Journal of Educational Psychology*, *Journal of Experimental Child Psychology*, *Journal of Learning Disabilities*, *Reading Research Quarterly*, and *Scientific Studies of Reading*. The studies discovered through the hand search, added to those found in the qualitative and quantitative reviews, totaled 41.

Authors of only 4 of these studies asked preschool teachers to train their children and conduct the experimental program (Byrne & Fielding-Barnsley, 1995; Haddock, 1976; Rosner, 1974; Whitehurst et al., 1994); authors of an additional 9 studies, reported in 8 articles, did so at kindergarten (Blachman et al., 1994; Brady, Fowler, Stone, & Winbury, 1994; Bus, 1986; Lundberg et al., 1988; O'Connor et al., 1996; Olofsson & Lundberg, 1983; Schneider et al., 1997; Tur-Kaspa, Mioduser, & Leitner, in press). Further, 1 of 4 preschool studies (Byrne & Fielding-Barnsley, 1995) and 6 of 9 kindergarten studies (Bus, 1986; Lundberg et al., 1988; Olofsson & Lundberg, 1983; Schneider et al., 1997; Tur-Kaspa et al., in press) were conducted in Australia, the Netherlands, Denmark, Sweden, Germany, or Israel; that is, only 6 of the 13 published phonological awareness training studies with teachers as trainers were conducted in English with American preschoolers and kindergartners.

This is important for at least two reasons. First, American children typically enter kindergarten 1 year earlier than do German children and 1 to 2 years before many Scandinavian children (see

Lundberg et al., 1988; Schneider et al., 1997). The typical German and Scandinavian kindergartner, therefore, has probably reached a relatively advanced level of cognitive development (Lundberg et al., 1988). Second, English has a less regular orthography than do several European languages such as Danish, Finnish, and German.

There are more concerns about the 13 studies. None of the researchers provided data on the fidelity (or accuracy) with which teachers implemented the training program; investigators of only 2 studies (Bus, 1986; Whitehurst et al., 1994) randomly assigned teachers (and their classrooms) to experimental and control groups; and authors of only 6 studies (Bus, 1986; Byrne & Fielding-Barnsley, 1995; Haddock, 1976; Olofsson & Lundberg, 1983; Rosner, 1974; Tur-Kaspa et al., in press) controlled for Hawthorne effects. Without fidelity information, one cannot attribute (positive or negative) findings to the treatment. Without use of random assignment and a means of controlling for possible Hawthorne effects, one cannot dismiss many nontreatment factors as responsible for study outcomes.

In sharing these observations, we do not wish to be harsh. We know the many difficulties involved in field-based research. We are respectful of the effort reflected in these 13 studies, and we understand that no field-based investigation is perfectly conceptualized and executed. Still, without well-controlled experiments showing that teacher-implemented programs strengthen children's reading readiness, the field may not rightfully claim that phonological awareness training has been validated as effective and feasible for classroom use.

Purpose of the Study

As indicated, many researchers who have implemented training studies have taught phonological awareness in isolation, rather than in combination with letter-sound or reading instruction. In some cases, such a strategy may reflect the cultural norms of study participants (see Lundberg et al., 1988). However, this approach also seems to reflect a belief that phonological awareness is a prerequisite of reading and that phonological awareness training should be conducted prior to, rather than concurrent with, formal reading instruction (e.g., Brennan & Ireson, 1997; Lundberg et al., 1988). Nonetheless, there is considerable support for combining the two (e.g., Ball & Blachman, 1991; Bryant & Bradley, 1985; Byrne & Fielding-Barnsley, 1993; Hatcher et al., 1994; Williams, 1980). Thus, further exploration of this issue is necessary, especially with teachers implementing treatments. Therefore, one of two treatment conditions in the present kindergarten study represents an integration of phonological awareness training with beginning decoding instruction and practice. We contrasted this group with (a) a treatment group that participated in phonological awareness training but not in decoding instruction and practice and (b) a no-treatment control group.

To draw relatively clear connections between these three groups and outcome data, we (a) randomly assigned kindergarten teachers (and their intact classes) within schools to the three study conditions; (b) collected treatment fidelity data on teachers and students; (c) observed teachers in all classrooms to describe their general literacy programs; (d) used well-regarded phonological awareness, beginning reading, and spelling measures; (e) required the phonological awareness and beginning reading treatments to run 20 weeks and 16 weeks, respectively; and (f) tested participants again

in the fall of the next year. To strengthen generalizability of our findings, we recruited a relatively large number of schools ($N = 8$), kindergarten teachers ($N = 33$), and children ($N = 404$). Half the schools were Title I (high-poverty) schools, and half served mostly middle-class children. In each classroom, we collected data on low-achieving, average-achieving, and high-achieving students, 25 of whom had disabilities.

Further, we explored the specificity of treatment effects, that is, (a) whether children who participate in phonological awareness training but not in decoding instruction and practice demonstrate improved performance on phonological awareness measures but not on reading or spelling measures and (b) whether students in both phonological awareness training and decoding instruction and practice show growth on phonological awareness, reading, and spelling measures. Although such results may seem self-evident on the basis of research showing that "you get what you teach," investigators have not frequently examined the specificity of treatment effects with respect to phonological awareness training and decoding instruction in kindergarten, especially when teachers have conducted the interventions.

Finally, this study represents an initial evaluation of Peer-Assisted Learning Strategies (PALS) in kindergarten. PALS is a classwide, peer-mediated approach to instruction and practice that accelerates reading achievement in Grades 2-5 (e.g., Fuchs, Fuchs, Mathes, & Simmons, 1997; Simmons, Fuchs, Fuchs, Hodge, & Mathes, 1994) and Grade 1 (Mathes, Howard, Allen, & Fuchs, 1998). On the basis of this prior work, we wondered whether PALS might strengthen kindergartners' early reading development. Nevertheless, we were uncertain as to whether students as young as 5 years can make meaningful use of peer-mediated strategies. So we examined students' (and teachers') fidelity of PALS implementation as well as their progress on reading-related outcomes.

Method

Treatments

"Ladders" was our phonological awareness training treatment. "Ladders + PALS" was our phonological awareness training and beginning decoding approach. Before describing the nature of these treatments, we provide some context by discussing the reading/language arts programs of the teachers in the study.

Reading/language arts context. On the basis of one-to-one interviews with all participating teachers, their completion of a questionnaire, and a 20-week presence in their classrooms, we concluded that multiformity, not uniformity, characterized what and how they taught. Virtually all teachers made use of whole-language activities. However, their focus and emphasis varied (e.g., journal writing vs. shared reading), as did the quality of their teaching. Nearly two thirds of our teacher sample used the school district's formally adopted text: *Harcourt-Brace Treasury of Literature: First Street Collection for Kindergarten* (Farr & Strickland, 1995). A majority of the teachers said they made frequent use of First Street's "Big Books." About half of the teachers reported using High Hat (Goldman & Lynch, 1986) animal stories and picture cards emphasizing letter-sound correspondence and blends. Only 1 (control) teacher systematically led her students through a phonics-based program of beginning reading. We failed to discern clear-cut differences in teachers' reading/language arts programs between schools or study groups. The one exception to this conclusion is that a majority of control teachers taught alphabet letter naming, whereas most treatment teachers did not. We do not know whether our treatments were

responsible for this difference. Finally, there were no statistically significant differences among the three study groups in the number of hours per week teachers reported devoting to language arts, $F(2, 30) = .00$, *ns*, or to reading, $F(2, 30) = .81$, *ns*. For language arts, the mean number of hours per week reported by the Ladders + PALS, Ladders, and control teachers was 11.46, 11.55, and 11.55, respectively; for reading, teachers reported 7.00 hours, 5.55 hours, and 7.64 hours, respectively.

Ladders. With Rollanda O'Connor's help, we chose 15 Ladders activities from more than 80 lessons in her *Ladders-To-Literacy* workbook (O'Connor, Notari-Syverson, & Vadasy, 1998b). These 15 activities were chosen for two reasons: to help teachers promote phonological awareness among their students and to differentiate Ladders from the beginning-decoding PALS component. Of the 15 activities, 10 were designed to stimulate word and syllable awareness, rhyming, first-sound isolation, onset-rime-level blending, and sound segmentation. Six of the 10 activities were chosen to promote the blending or segmenting of sounds in consonant-vowel-consonant words; only 1 activity required manipulations of printed letters. Teachers conducted the 10 activities three or more times during 2 nonconsecutive weeks. The remaining 5 Ladders activities were journal writing, "letter sound of the week," "morning message," nursery rhymes and poems, and shared storybook reading. Only journal writing and morning message presented students with printed letters. Each of these 5 activities was conducted at least once per week for the entire implementation period. All 15 Ladders activities were teacher led, were directed to the whole class, and required 5 to 15 min each day of implementation. The maximum time teachers devoted to Ladders each week was 45 min (15 min \times 3 days), or 10% of their reading/language arts program. Teachers conducted Ladders for 20 weeks.

O'Connor, Notari-Syverson, and Vadasy (1996, 1998a) reported that Ladders improved the early literacy skills of special-needs students and nondisabled low-income children. However, O'Connor et al. used 25 activities, in contrast to our 15, and their activities deliberately placed greater emphasis on three-phoneme segmentation and blending tasks.

Ladders + PALS. The PALS component of the more complex Ladders + PALS treatment required children to work in dyads with same-age peers on as many as 51 PALS lessons. With the Rapid Letter Naming (RLN) Test, a predictor of future reading performance (e.g., Levy & Stewart, 1991; O'Connor, Jenkins, & Slocum, 1995; Torgesen, Wagner, & Rashotte, 1997; Wolf, 1991), we paired the highest scoring student in each class with the lowest scoring student, the second-highest scoring student with the next-to-last scoring student, and so on. If a teacher believed a pair to be socially incompatible, the pair was reassigned. Each student in each pair took a turn as reader (tutee) and coach (tutor). Pairs remained together for 4 to 6 weeks, at which point the teacher named new pairs. Teachers trained their students to work productively and cooperatively. Before each PALS lesson, for about 5 min, the teachers modeled new letter sounds and sight words, increasing the likelihood that their students would experience success. Previous research indicates that PALS strengthens the word attack, word identification, reading fluency, or reading comprehension of first-grade students (Mathes et al., 1998) and second through fifth-grade students (e.g., Fuchs et al., 1997; Simmons et al., 1994). Teachers conducted PALS three times per week for 16 weeks. Sessions lasted about 20 min beyond the brief teacher-led instruction. This represented 10% to 15% of teachers' reading/language arts time. Thus, the combined Ladders + PALS treatment accounted for a maximum of 20% to 25% of teachers' reading/language arts program. This was more than double the time of the Ladders treatment. Nevertheless, as already documented, the total amount of reading/language arts time for students in the two treatment groups was virtually identical.

In kindergarten PALS, the What Sound? activity (WS) is the first of two activities. Its purpose is to help students learn the correct sounds of all 26 letters except *x*. In this activity, the coach points to a printed letter and asks, "What sound?" The reader responds, after which the coach provides praise for an appropriate answer or a standard correction for an incorrect answer.

The coach's correction is as follows: "Stop. You missed that sound. That sound is [letter sound]." The coach again asks, "What sound?" After a correct response, the coach says, "Good. Read that line [of letters] again." There are 51 WS lessons printed on separate, 9-in. \times 11-in. sheets of paper. Each sheet contains four lines of upper- and lowercase letters, with six letters per line. Interspersed among the letters are prominent black stars. When the pair gets to a star, the coach tells the reader, "Good job!" A new letter sound is introduced in approximately every other lesson. The order in which letters are introduced in the WS lessons, as well as the procedure and layout, of the lessons, was influenced by Carnine, Silbert, and Kameenui (1990).

The second PALS activity is called What Word? (WW). It requires children to read aloud sight words, decodable words, and simple sentences. Reading orally permits a child's partner and teacher to monitor word identification skill and to present corrective feedback when errors are made. The sight words are *I, the, is, was, on, and has*. Teachers introduce the first of them in Lesson 29 and new ones at an approximate rate of one every three lessons. On the same lesson sheet containing the WS activity, the coach points to both newly presented and previously learned sight words and asks, "What word?" The correction procedure for this activity is similar to the one used in the WS activity.

Also on the same lesson sheet are decodable words (i.e., words that can be sounded out with letter sounds practiced in earlier lessons) representing as many as five word families: *at, an, ap, ad, and am*. Each letter of each decodable word is placed in a "sound box" (Elkonin, 1973). The coach says, "Read the word slowly." The reader slowly says the letter sounds as he or she touches the individual sound boxes. Then the coach asks, "What word?" Simple sentences (still on the same lesson sheet) are composed of the sight words and decodable words learned in previous WW lessons. When the reader misreads a sentence, the coach applies a correction procedure similar to those used in connection with the reading of sight words and decodable words in the WS activity. Teachers implementing the Ladders + PALS treatment were trained and expected to conduct the Ladders lessons with the same frequency and degree of integrity as did the Ladders-only teachers.

Control. Prior to the study, control teachers were told they were members of a control group, and we discussed with them (and with the teachers in the Ladders and Ladders + PALS treatments) why such a group was important in evaluating the effectiveness of Ladders and PALS. We asked control teachers to continue their reading/language arts instruction and not to conduct Ladders or PALS lessons for the study's duration. We promised that, following completion of the study, we would give them copies of Ladders and PALS manuals. Such promises and discussion were deemed necessary because teachers within a given school had been assigned randomly to the three study groups, with the result that many control classrooms were literally next door to Ladders and Ladders + PALS classrooms.

School, Teacher, and Student Selection

Schools. We recruited four Title I and four non-Title I schools in the Metro-Nashville Public Schools system. Douglas Fuchs and Anneke Thompson discussed the study with principals and teachers of 10 schools; only 1 school declined participation. Of the remaining 9, we chose the first 8 that agreed to participate. Median percentages of African American kindergarten children in the Title I and non-Title I schools were 52% and 21%, respectively. For the study classrooms only, the median proportion of children receiving free or reduced lunch was 81% in Title I schools and 29% in non-Title I schools. The average price of a home in the communities surrounding the four Title I schools was \$53,923, versus \$136,309 in the non-Title I school communities.

Teachers. Douglas Fuchs and Anneke Thompson explained to the teachers the study's rationale and purpose, the general nature of the Ladders and Ladders + PALS treatments, and that we wished to randomly assign them to treatment conditions. Among the eight study schools, 38

of 43 kindergarten teachers volunteered to participate. We chose 33, ensuring that 18 taught in Title I schools and that 15 taught in non-Title I schools. Title I teachers were overselected (i.e., 1 extra teacher for each of the three study groups) to compensate for an expected greater student turnover in those schools.

The 33 teachers were assigned to study groups by means of stratified randomization. Specifically, teachers in each of the eight schools were assigned randomly to the three study groups. Title I teachers ($n = 18$) and non-Title I teachers ($n = 15$) were distributed equally among the three groups such that 6 and 5, respectively, were assigned to each, totaling 11 teachers per group. We conducted one-way analyses of variance (ANOVAs; Ladders + PALS vs. Ladders vs. controls) and found no significant differences among groups on teachers' class size or years of teaching experience. We ran chi-square analyses and did not obtain reliable relations between study groups and teachers' age, gender, highest degree earned, or race (see Table 1).

Students. Two criteria guided the selection of students as study participants: the RLN test and teacher judgment. Project staff individually administered the RLN test to all students in participating classrooms. The RLN test is a timed measure of alphabet letter names, and as mentioned earlier, research indicates that it is a respectable predictor of future reading performance. Children with the lowest six scores in each of the 33 classrooms were designated by staff as low achievers (LA). Four students with scores in the middle of the distribution were assigned average-achiever (AA) status. The four students with the highest scores were labeled high achievers (HA). LAs were overselected because, in a pilot study, these students proved most likely to leave before school was out.

The names of the 6 LA, 4 AA, and 4 HA students were shown to each teacher for validation. When a teacher did not concur with an LA, AA, or HA rating, the child in question was replaced by a student with the next closest RLN score, and his or her name was then presented to the teacher for approval. In this manner, 379 children were assigned to an LA, AA, or HA category. Across the 33 study classes, an additional 25 children were identified as special-education students on the basis of a current individualized educational plan (IEP). They, too, were rated LA, AA, or HA, resulting in a final study sample of 404 children.

Two-factor ANOVAs (treatment: Ladders + PALS vs. Ladders vs. controls; student type: LA vs. AA vs. HA) on number of school absences, age, and report card grades indicated students were comparable across study group. Further, there were no significant interactions between study group and student type. However, two significant main effects were identified for student type. Follow-up tests revealed that (a) HA students had fewer school absences than did LA and AA students, whereas LA and AA students were similar in this regard; and (b) LA students had lower report card grades than did AA or HA students whose grades were comparable. In addition, chi-square tests indicated no reliable relations between the study groups and students' Title I, English-as-a-second-language (ESL), special-education, or retention status, or between study groups and students' race or gender (see Table 2).

Staff and Classroom-Based Assistance

The four project staff were women. One worked 40 hr per week, and three worked 20 hr per week. Two were special-education graduate students, a third was a graduate student in developmental psychology, and the fourth had a master's degree in special education. Three were experienced classroom teachers. The number of study teachers assigned to each staff person ranged from 5 to 11 (median = 8.50).

Across the three study groups, staff administered pre- and posttreatment tests and follow-up tests, collected demographic data on students and teachers, and conducted structured interviews with teachers regarding their reading/language arts programs. Staff also gathered data on how frequently, and with what degree of fidelity, Ladders and Ladders + PALS teachers implemented treatment activities. In addition, they answered teachers' questions and occasionally offered suggestions about how best to implement the treatments. In Ladders + PALS classes, staff helped teachers train their students in PALS. Staff visited each Ladders and Ladders + PALS classroom twice each week for the first 2 months of treatment. Afterward, visits were reduced to once a week. A visit usually lasted about 40 min.

Table 1
Teacher Demographic Data by Study Group

Variable	Ladders + PALS ($n = 11$)			Ladders ($n = 11$)			Control ($n = 11$)			F^a	$\chi^2(df)$
	<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> (%)		
Age											3.47 (6)
21-30			3 (27)			1 (9)			2 (18)		
31-40			1 (9)			3 (27)			1 (9)		
41-50			5 (45)			5 (45)			7 (64)		
51+			2 (18)			2 (18)			1 (9)		
Class size (No. of pupils)	18.55	1.81		19.00	1.61		19.18	2.36		.31	
Female teachers			11 (100)			11 (100)			11 (100)		
Highest degree earned											.79 (4)
B.S./B.A.			5 (45)			4 (36)			3 (27)		
M.Ed./M.S.			6 (55)			7 (64)			8 (73)		
Ed.S/Ph.D			0 (0)			0 (0)			0 (0)		
Race											1.50 (4)
African American			1 (9)			1 (9)			2 (18)		
Caucasian			9 (82)			9 (82)			9 (82)		
Other			1 (9)			1 (9)			0 (0)		
Teaching experience (in years)	15.82	9.96		15.18	8.64		15.00	8.80		.02	

Note. Because of rounding, percentages may not sum to 100. PALS = Peer-Assisted Learning Strategies.
^a $df = 2, 30$.

Table 2
Student Demographic Data by Study Group and Student Type

Variable and student type	Ladders + PALS (n = 133)			Ladders (n = 136)			Control (n = 135)			F(df) ^a	F(df) ^b	F(df) ^c	χ^2 (df)
	M	SD	n(%)	M	SD	n(%)	M	SD	n(%)				
Absences ^d										.91 (2, 395)	6.66* (2, 395)	0.99 (4, 395)	
LA	1.96	2.31		2.33	2.47		2.88	3.88					
AA	2.59	2.58		1.69	2.08		2.05	3.83					
HA	1.28	1.80		1.02	1.46		1.45	2.19					
Chronological age										.50 (2, 379)	1.13 (2,379)	1.14 (4, 379)	
LA	5.65	.48		5.55	.35		5.61	.38					
AA	5.66	.35		5.80	.45		5.60	.35					
HA	5.61	.94		5.72	.44		5.68	.37					
ESL ^e													
LA			4 (3)			3 (2)							.20 (2)
AA			2 (2)			1 (1)							.33 (1)
HA			2 (2)			0 (0)							.33 (1)
IEP classification													
Language/speech													
LA			7 (5)			5 (4)							.88 (2)
AA			0 (0)			0 (0)							.3 (2)
HA			2 (2)			0 (0)							.33 (1)
Other													
LA			0 (0)			0 (0)							1 (1)
AA			0 (0)			1 (1)							0 (0)
HA			0 (0)			0 (0)							1 (1)
Kindergarten report card ^f										.94 (2, 349)	8.89** (2, 349)	.50 (4, 349)	
LA	36.50	6.26		35.87	5.58		35.44	5.51					
AA	38.72	1.75		38.56	2.05		37.21	3.38					
HA	37.75	7.22		38.89	2.40		38.18	6.56					
Race													
African American													
LA			14 (11)			18 (13)							.64 (2)
AA			19 (14)			19 (14)							.60 (2)
HA			17 (13)			18 (13)							.11 (2)
Caucasian													
LA			31 (23)			33 (24)							.24 (2)
AA			16 (12)			17 (13)							.49 (2)
HA			24 (18)			18 (13)							.90 (2)
Other													
LA			6 (5)			6 (5)							.12 (2)
AA			4 (3)			3 (2)							.67 (2)
HA			2 (2)			4 (3)							2.00 (2)
Gender (female)													
LA			18 (14)			21 (15)							.43 (2)
AA			22 (17)			15 (11)							.44 (2)
HA			25 (19)			23 (17)							.08 (2)
Title I ^g													
LA			28 (21)			32 (24)							.36 (2)
AA			22 (17)			20 (15)							.23 (2)
HA			24 (18)			22 (16)							.21 (2)

Note. PALS = Peer-Assisted Learning Strategies; LA = low achievers; AA = average achievers; HA = high achievers.

^a F values for treatment main effect.

^b F values for student-type main effect.

^c F values for Treatment \times Student-Type interaction.

^d Number of absences in the fourth 6-week period of the school year.

^e Number of children who were in the English-as-a-second-language program.

^f Kindergarten report card score during the first 6-week period of the school year. Scores range from 0 to 40.

^g Title I status refers to school, not to individual students. For example, the number 28 associated with LA students in the Ladders + PALS group indicates that 28 such children attended a Title I school. We do not know whether each (or any) of these children received free or reduced lunch.

* $p < .01$. ** $p < .001$.

Measures

RLN. We have discussed how RLN was used to identify student participants as LA, AA, and HA. It was also used as a pretreatment

variable. RLN assesses the number of letter names a student identifies in 1 min. Students are given a 9-in. \times 11-in. sheet displaying 52 letters. Each letter of the alphabet is displayed in upper- and lowercase, on seven lines, in random order. Uncovering one line at a time, the examiner points to a

letter and says, "Go across the page and tell me the names of as many letters as you can. Try to name each letter. If you come to a letter you don't know, I'll tell it to you. It's OK if you don't know the names of many of these letters. The important thing is to try your best." If the student does not respond in 3 s, the examiner says the name of the letter and moves to the next. The number of letters named correctly in 1 min is the student's score. If the student completes the test in less time, the score is prorated.

Ladders fidelity. To determine how often Ladders activities were used, monthly calendars were distributed to the teachers. The calendars offered a reasonable implementation sequence and timeline. The teachers recorded the activities they implemented and when they did them.

To explore the quality of Ladders implementation, staff assigned teachers a weekly global rating ranging from 1 (*poor*) to 3 (*excellent*). The ratings were designed to reflect (a) lesson clarity; (b) how well the teacher's presentation jibed with the intent of the lesson; and (c) the degree to which the students—LA and special-education students in particular—were engaged. Before these ratings were conducted, staff simultaneously coded videotapes of three teachers conducting the same activity with varying levels of fidelity. Complete (100%) agreement was obtained among all staff on these tapes before the treatment fidelity ratings commenced in classrooms.

PALS fidelity. On two occasions, staff used checklists to evaluate the accuracy with which teachers and students used PALS. The checklist for the first occasion was developed for the WS activity. It consists of 12 teacher behaviors and 64 student behaviors. Teacher behaviors address how they begin and end the activity and how they monitor the lesson. Student behaviors address purposefulness, organization, and accuracy of implementation.

The second PALS checklist evaluates implementation of both WS and WW activities. It comprises 13 teacher behaviors and 214 student behaviors. The staff observed three randomly chosen pairs during a PALS lesson. Pairs with an ESL student, a special-education student, or a student new to the PALS program were not observed. Observers rotated from one pair to another after each partner had been observed in the roles of reader and coach. On both PALS checklists, behavior is scored as "demonstrated," "not demonstrated," or "not applicable." Each observation yields an overall teacher score, an overall (averaged) student score, and a combined teacher and student classroom score. Scores are derived by dividing the number of behaviors demonstrated by the number of behaviors demonstrated and not demonstrated and multiplying by 100. For both the first and second fidelity checks, two staff members simultaneously observed teachers and students in four classes (2 in Title I schools and 2 in non-Title I schools) to determine interrater agreement. For the first fidelity check, agreement was 100%, 99%, and 99% for teacher, student, and teacher-student combined, respectively. For the second check, respective data were 100%, 100%, and 98%.

Rapid Letter Sound (RLS). The RLS test, which is based on a measure developed by Levy and Lysunchuk (1997), assesses the number of letter sounds a student identifies in 1 min. The student is given a 9-in. \times 11-in. sheet that provides four practice letters and displays, in random order, all 26 lowercase letters on five lines. The examiner says, "I'm going to show you some letters. You tell me what sound the letters make. If you don't know the sound a letter makes, don't worry. What's important is that you try your best." The examiner then shows the student only the four practice letters and says, "This letter says /b/. Your turn. What sound does it say?" After providing the student with practice, the examiner uncovers the first line of the test proper and says, "You're doing a great job. Now it's just going to be your turn. Go as quickly and carefully as you can. Remember to tell me the sounds the letters make. Try your best. If you don't know a letter sound, it's okay." If the student fails to give a response in 3 s, the examiner moves to the next letter. After 1 min, the number of sounds expressed correctly is the score. If a student completes the test prior to 1 min, the score is prorated. The RLS test was administered at pre- and posttreatment and at follow up.

Segmentation. This timed, 1-min measure closely resembles the Yopp-Singer Test (cf. Yopp, 1988) and assesses children's ability to deconstruct words into component sounds. It consists of 3 three-phoneme practice words (e.g., *dog*) and 22 two- or three-phoneme words. The examiner uses the practice words to help the student understand the task. If the student cannot produce a single correct sound in at least one practice word, the test is stopped. For the remaining 22 words, the examiner says, "Say the sounds in [word]." If the student does not say one correct sound in 4 consecutive words, the test is terminated. The student's score is the number of correct phonemes expressed in 1 min. The segmentation test was given before and after treatment and at follow up.

Word Attack Subtest of the Woodcock Reading Mastery Test—Revised, Form G (Woodcock, 1987). This measure evaluates students' ability to pronounce pseudowords. It is commonly used (e.g., Torgesen et al., 1997) because it is widely regarded as a sensitive test of decoding skill and reading progress (e.g., Felton & Wood, 1992; Juel et al., 1986; Tunmer & Hoover, 1992). The Word Attack subtest contains 45 nonsense words, progressing in presentation from most easy to most difficult. The test is discontinued after six consecutive errors. Students earn 1 point for each correctly pronounced word. Scores range from 0 to 45. Split-half and test-retest reliabilities reported in the manual are .95 and .90, respectively, for first grade. No such information was given for the kindergarten children in the standardization population. This subtest was given at pretreatment, at posttreatment, and at follow up.

Word Identification Subtest (Word ID) of the Woodcock Reading Mastery Test—Revised, Form G (Woodcock, 1987). The Word ID subtest is a frequently used measure of real-word reading ability (Vellutino & Scanlon, 1987; Torgesen et al., 1997). It requires children to read single words out of context. It consists of 100 words. For each child, we started with Item 1 and discontinued testing after six consecutive errors. The subtest was administered and scored in standard fashion. Students earn 1 point for each correctly pronounced word. Scores range from 0 to 100. Split-half and test-retest reliabilities are .99 and .94, respectively, for first grade; again, no reliability or validity data were found for kindergarten children. The subtest was administered both before and after treatment implementation and at follow up.

Blending. We created a test consisting of 22 three-sound words (e.g., *soap, mom, food*). On four practice items preceding the test proper, the examiner says, "I'm going to say some sounds. If you put the sounds together, they make a word." Then the examiner says, "/C/ /a/ /t/. What word is that?" If the student commits four consecutive errors, the test is stopped. Otherwise, the examiner records the number of words blended correctly in 1 min and continues testing until all 22 test items are presented. The number of words identified correctly in 1 min is the student's blending score, which ranges from 0 to 22. The blending test was administered at posttreatment and at follow-up.

Spelling Subtest of the Wechsler Individual Achievement Test. The Spelling subtest consists of 50 words. To obtain a sufficient sample of spelling behavior, we administered to every child the first 12 test items (Items 1 to 6 address letter recognition; Items 7 to 12 ask children to spell single words). Otherwise, the test was administered in standard fashion: For each item, the examiner said the word, then said the word in a sentence, then repeated the word. Students had 10 s to write the word. Scoring was conducted in accordance with the developmental scoring rubric created by Tangel and Blachman (1992). We used a developmental score because of our study participants' age: Whereas many kindergartners are incapable of traditional spelling, many can write the first letter of a word, or they may spell a word phonetically. Accordingly, Items 1 to 6 were scored in standard fashion; but on Items 7 to 50, students could receive a score of 0 to 6, resulting in a total Spelling score ranging from 0 to 270. We created explicit scoring criteria for Items 7 to 50, reflecting Tangel and Blachman's (1992) guidelines. Three staff members independently scored 20% of the protocols, achieving 94% agreement. After scoring was completed by these staff members, one member rescored 10% of the protocols scored by

others; agreement was 96%. The Spelling subtest was administered at posttreatment and at follow-up.

PALS teacher questionnaire. PALS teachers completed a posttreatment questionnaire, which asked them to evaluate the effectiveness of PALS on a 5-point Likert-type scale and to explain how they regularly found time to implement the activities. Although the teachers were asked to identify themselves on the questionnaires, we told them repeatedly (when we handed them the instrument and on the questionnaire itself) that we were interested in their candid appraisals and opinions and that their individual responses would not be shared with anyone other than project staff.

Procedure

Training. Teachers attended a full-day workshop during which we discussed phonological awareness in terms of blending sounds into words, segmenting words into sounds, and rhyming words to hear similarities of sound. We also discussed the connections between sounds and printed letters and between letters and words, and we described the 15 Ladders activities teachers would be asked to implement. We explained that some Ladders activities are relatively easy (e.g., Guess My Word; word sounds are stretched in exaggerated fashion [*mmaasskk*]), several Ladders activities are more challenging (e.g., I'm Thinking Of; requires awareness of onset-rime [*m-ask*]), and a few activities are more difficult still (e.g., segmenting words at the phonemic level [*m-a-s-k*]). We demonstrated how, irrespective of difficulty, all Ladders activities can be presented with varying degrees of teacher support, permitting teachers to use each with virtually all students in the class. We also emphasized the importance of involving LA children as well as HA children, maintaining a brisk-paced lesson, and cycling through the activities to promote a sense of freshness and novelty. Finally, each teacher was given a Ladders manual, and staff members and teachers carefully reviewed its content.

Ladders + PALS teachers attended an additional half-day workshop to prepare them to train their students in PALS. Teachers' logistical concerns were addressed by providing suggestions on how to pair students, assign seats, and schedule activities. A proposed time line for PALS implementation was discussed, and teachers were given detailed examples of typical lessons. To promote familiarity with these lessons, teachers formed dyads and, under the direction of the staff, role played each activity as both coach and reader. Teachers were encouraged to frequently model for their students cooperative and supportive PALS behavior and, during PALS lessons, to walk about the classroom, giving praise to student pairs who were working collegially. Finally, teachers were given a comprehensive manual, written expressly for them, that included scripted PALS lessons and acetate overheads to facilitate student training. (We encouraged teachers to translate the scripts into their own words.)

Fidelity. Fidelity-of-treatment observations occurred weekly in each Ladders classroom between Week 6 and Week 20, inclusive. PALS treatment fidelity data were collected in each classroom on two occasions. In Week 9, staff members observed teachers' and students' use of the WS activity; in Week 16, staff members observed teachers' and students' implementation of both WS and WW activities. Guiding these observations was a checklist reflecting important PALS behaviors for teachers and students. A description of the checklist and related information is in the *Measures* section.

Testing and scoring. All tests were administered to students on a one-to-one basis. Testing in kindergarten (pre- and posttreatment) was accomplished in two sessions. The RLN test was administered first at pretreatment. Other measures were given in random order across children and pre- and posttreatment testing. Following completion of the treatments, during the second testing session, the blending test was always administered before the spelling test. The staff was trained to administer the tests according to instructions in a user's manual. In the absence of such instructions, they practiced agreed-on procedures to ensure standard administration. Because the staff tested children in classrooms in which they

provided ongoing assistance, they were familiar to the children. Because evidence suggests examiner unfamiliarity may depress the performance of young students from low-income families (e.g., Fuchs & Fuchs, 1986) and children with disabilities (e.g., Fuchs, Fuchs, Power, & Dailey, 1985), staff members tried to make themselves familiar to control children by spending between 1 and 5 hr in their classes a couple of weeks before posttreatment testing.

A related potential problem was that staff members knew children's assignments to the three study groups. Staff members were told repeatedly that the purpose of the study was not to prove that one treatment was better than another or that both were superior to controls. Rather, it was simply to conduct every phase of the study with care and then "let the chips fall as they may." For example, before posttreatment testing (spring of the kindergarten year) and follow-up testing (fall of the next school year), staff members were told that their responsibility was to administer the measures in standard fashion and to try to elicit the optimal performance of all children. We discussed research on experimenter bias (e.g., Rosenthal, 1980) and the various ways it can undermine otherwise careful evaluations.

All measures were scored by the test administrator and rescored by a second staff member. The one exception to this was the developmental scoring rubric used for the Spelling test. All scores were entered into a computer database by two staff members working collaboratively. After data entry, the same two staff members systematically examined the data set for transpositional errors.

Among the 404 children we tested in fall and spring of their kindergarten year, we found 312 to test again the following October. Of these 312 children, 294 were enrolled in 74 first-grade classes, and 18 were repeating kindergarten in 15 classes. We found our follow-up sample in 89 classrooms in 18 schools. Measures, testers, and testing procedures were virtually the same at follow-up as they were in the previous year. One exception was that whereas testing was accomplished in two sessions in the prior year, it was completed in one session at follow up. This change was dictated partly by scheduling difficulties. However, the children were older and, we believed, more capable of maintaining attention and effort for a longer duration.

Results

Fidelity of Treatment

Ladders. Information on Ladders implementation came from three sources: the calendars on which teachers recorded their use of the activities, our direct observations of their lessons, and the teacher questionnaire. Teachers' calendar notes indicated that they typically spent between 5 and 15 min on each lesson, and most teachers conducted them with the frequency and in the sequence that we had recommended. Questionnaire responses suggested that the relationship between teachers' use of Ladders and their overall reading/language arts program depended on whether teachers used Ladders alone or in combination with PALS. As shown in Table 3, 7 of 11 Ladders teachers increased their reading/language arts time to accommodate Ladders. In contrast, only 1 of 11 Ladders + PALS teachers increased their reading/language arts time, with 10 choosing to implement the lessons by supplanting something else. In other words, a majority of Ladders teachers used the lessons as a supplement, whereas virtually all Ladders + PALS teachers used them as a substitute. Finally, our 3-point quality ratings of Ladders lessons, averaged across teachers and weeks, was 2.28 ($SD = .27$) and 2.25 ($SD = .21$) for Ladders teachers and Ladders + PALS teachers, respectively.

PALS. As mentioned, the accuracy of PALS implementation was evaluated twice. At Time 1 (Week 9 of the treatment), teach-

Table 3
Influence of Ladders and PALS Implementations on Teachers' Reading/Language Arts Programs

Teachers, treatments, and questions	No. of teachers responding "yes"
Ladders teachers/Ladders treatment	
1. Did Ladders replace all teacher-directed instruction?	0
2. Did Ladders replace part of teacher-directed instruction?	3
3. Did Ladders replace independent seatwork?	0
4. Did Ladders replace the entire reading/language arts program?	1
5. Was Ladders conducted at another time?	7
Ladders + PALS teachers/Ladders treatment	
1. Did Ladders replace all teacher-directed instruction?	0
2. Did Ladders replace part of teacher-directed instruction?	10
3. Did Ladders replace independent seatwork?	0
4. Did Ladders replace the entire reading/language arts program?	0
5. Was Ladders conducted at another time?	1
Ladders + PALS teachers/PALS treatment	
1. Did PALS replace all teacher-directed instruction?	0
2. Did PALS replace part of teacher-directed instruction?	6
3. Did PALS replace independent seatwork?	1
4. Did PALS replace the entire reading/language arts program?	0
5. Was PALS conducted at another time?	4

Note. PALS = Peer-Assisted Learning Strategies.

ers' and students' use of the WS activity was observed; at Time 2 (Week 16), their use of the WS and WW activities was evaluated. Teachers' averaged accuracy of implementation was 85% ($SD = 5.80$) at Time 1 and 82% ($SD = 11.84$) at Time 2. Student accuracy was 87% ($SD = 10.76$) and 77% ($SD = 12.36$) at Times 1 and 2, respectively. Overall (teacher and student combined) accuracy of implementation was 87% ($SD = 8.90$) at Time 1 and 77% ($SD = 11.68$) at Time 2.

Student Performance in Kindergarten and at Follow-Up

In analyzing student performance data, we used teacher as the unit of analysis because teachers, rather than individual students, had been assigned randomly within schools to the study groups. This required us to average the scores of each teacher for all three (LA, AA, and HA) student groups.

To examine specificity of treatment effects, we used multivariate analyses of variance, combining conceptually similar and highly correlated measures into two constructs: *phonological awareness* and *alphabetsics*. Phonological awareness subsumed segmenting and blending tasks; alphabetsics included RLN, RLS, Word ID, Word Attack, and Spelling measures. Use of these two constructs corresponded with our most basic questions about the treatments: Does phonological awareness training differentially affect student performance on phonological tasks? and Does beginning decoding instruction and practice influence student performance on letter sounds, reading, and spelling measures?

To explore whether the deliberate variations in our treatments differentially affected student performance, we stipulated two Helmert contrasts (Stevens, 1986) for the phonological awareness construct and two for the alphabetsics construct. Regarding phonological awareness, with the first multivariate contrast, we explored whether students in the Ladders and Ladders + PALS treatments (both of which incorporated phonological awareness training) outperformed controls; with the second contrast, we asked whether the Ladders and Ladders + PALS groups differed. For alphabetsics,

with the first contrast, we addressed whether students in the Ladders + PALS treatment (the only treatment that focused systematically on beginning decoding) outperformed students in the other two groups; with the second contrast, we looked at whether the Ladders and control groups performed differently from each other.

Because our preliminary analyses indicated that a school's Title I status did not mediate treatment effects on either the phonological awareness or alphabetic constructs, we excluded Title I status from subsequent analyses. In addition, because we found no pretreatment differences among study groups, we conducted the multivariate contrasts on posttreatment performance and follow-up measures.

To calculate the magnitude of multivariate effects, we used eta squared (for which .01 is small, .06 is medium, and .14 is large, according to Cohen [1977]). To derive effect sizes that could be compared with the Bus and van IJzendoorn (1999) meta-analysis, we computed Cohen's d by dividing the difference between treatment means by the pooled standard deviation on the univariate posttreatment (and follow-up) variables (Hedges & Olkin, 1985). Table 4 shows the pre- and posttreatment scores of the full sample of 404 children. Table 5 reports pretreatment, posttreatment, and follow-up scores for the 312 children we located in the fall of the next school year.

Pretreatment. No significant pretreatment effects were identified. On phonological awareness (segmentation), for LA, AA, and HA students, respectively, the $F(1, 30)$ for Contrast 1 (Ladders and Ladders + PALS vs. control) was .43, 1.27, and .10; for Contrast 2 (Ladders vs. Ladders + PALS), it was .58, 1.66, and .15, respectively. On alphabetsics (RLN, RLS, Word Attack, and Word ID), for LA, AA, and HA students, respectively, the $F(4, 27)$ for Contrast 1 (Ladders + PALS vs. Ladders and control) was 1.25, 2.19, and .47; for Contrast 2 (Ladders vs. control), it was .92, .36, and .20, respectively.

Table 4
Kindergarten Reading Performance by Student Type and Study Group

Compound variable, trial, and measure	LA						AA					
	Ladders + PALS		Ladders		Control		Ladders + PALS		Ladders		Control	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Phonological awareness												
Pretreatment												
Segmentation	1.25	1.26	1.73	2.11	1.97	1.63	4.49	4.79	2.87	2.76	2.08	1.84
Posttreatment												
Segmentation	11.83	6.79	11.72	6.71	8.39	7.67	19.60	5.31	15.76	6.73	7.99	5.74
Blending	6.90	4.10	6.40	4.16	2.48	2.85	10.47	3.85	7.91	3.56	3.52	3.18
Alphabetics												
Pretreatment												
RLN	5.29	4.27	5.09	3.88	4.33	4.11	17.16	8.27	18.87	8.62	16.08	7.89
RLS	2.32	2.08	1.26	2.08	1.77	2.11	7.51	6.23	4.94	6.05	4.07	3.85
WAT	0.04	0.15	0.00	0.00	0.06	0.15	0.33	1.00	0.86	1.02	0.22	0.47
Word ID	0.02	0.07	0.09	0.13	0.16	0.31	0.62	0.58	1.49	1.23	0.49	0.85
Posttreatment												
RLS	18.54	9.93	9.38	6.77	8.49	5.72	24.62	6.62	10.85	7.46	14.28	7.89
WAT	1.59	2.05	0.55	0.98	0.41	1.00	4.30	3.16	3.11	4.30	1.38	2.02
Word ID	2.21	3.32	1.27	1.28	1.00	1.34	6.81	6.14	6.63	8.47	2.87	4.59
Spelling	23.08	12.72	17.94	8.39	19.51	13.10	42.20	17.17	33.04	20.40	23.80	11.82

Note. LA = low achievers; AA = average achievers; HA = high achievers; PALS = Peer-Assisted Learning Strategies; RLN = Rapid Letter Naming;

Posttreatment. On phonological awareness (segmentation and blending), for LA, AA, and HA students, respectively, the $F(2, 29)$ for Contrast 1 (Ladders and Ladders + PALS vs. control) was 5.73 ($p < .01$), 10.86 ($p < .001$), and 5.40 ($p < .01$); for Contrast 2 (Ladders vs. Ladders + PALS), it was .10, 1.49, and .26 (all *ns*), respectively. Eta squared for the LA, AA, and HA students was .16, .27, and .15, respectively. For Ladders versus control, Cohen's d for segmentation and blending, respectively, was .46 and 1.12 for LA students, 1.25 and 1.30 for AA students, and 1.19 and 1.15 for HA students. For Ladders + PALS versus control, Cohen's d for segmentation and blending, respectively, was .45 and 1.27 for LA students, 2.10 and 1.97 for AA students, and 1.07 and 1.21 for HA students. For Ladders + PALS versus Ladders, Cohen's d for segmentation and blending, respectively, was .02 and .12 for LA students, .64 and .69 for AA students, and $-.15$ and .04 for HA students.

On alphabetics (RLS, Word Attack, Word ID, and Spelling), for LA, AA, and HA students, respectively, the $F(4, 27)$ for Contrast 1 (Ladders + PALS vs. Ladders and control) was 6.85 ($p < .001$), 2.75 ($p < .05$), and 2.38 ($p = .07$); for Contrast 2 (Ladders vs. control), it was .31, .45, and 1.57 (all *ns*). Eta squared for the LA, AA, and HA students, respectively, was .31, .18, and .10. For Ladders versus control, Cohen's d for the univariate variables ranged from .14 to .33 for LA students, $-.45$ to .58 for AA students, and .06 to .80 for HA students. For Ladders + PALS versus control, Cohen's d ranged from .28 and 1.28 for LA students, .73 to 1.42 for AA students, and .08 to .90 for HA students. For Ladders + PALS versus Ladders, Cohen's d for the univariate variables ranged from .41 to 1.10 for LA students, .02 to 1.96 for AA students, and .16 to .90 for HA students.

Normal Curve Equivalents (NCEs) were calculated for the study groups' posttreatment performance on Word ID and Word Attack. On Word ID, the Ladders + PALS group's average NCE was 38.80 ($SD = 15.30$) for LA students, 53.89 ($SD = 16.98$) for

AA students, and 67.11 ($SD = 21.12$) for HA students. Corresponding scores for the Ladders group were 32.83 ($SD = 16.19$), 43.68 ($SD = 25.51$), and 61.87 ($SD = 20.43$). Corresponding control group scores were 29.54 ($SD = 12.99$), 37.09 (20.88), and 61.38 ($SD = 16.20$). On Word Attack, the mean NCEs for Ladders + PALS LA, AA, and HA students were 47.39 ($SD = 13.74$), 59.65 ($SD = 14.80$), and 67.03 ($SD = 16.02$). Corresponding scores for the Ladders group were 42.00 ($SD = 11.26$), 47.74 ($SD = 21.79$), and 62.68 ($SD = 16.36$). Corresponding control group scores were 40.75 ($SD = 9.44$), 46.82 ($SD = 14.73$), and 54.49 ($SD = 19.54$).

Follow-up. On phonological awareness (segmentation and blending), for LA, AA, and HA students, respectively, the $F(2, 29)$ for Contrast 1 (Ladders and Ladders + PALS vs. control) was 3.40 ($p < .05$), 4.78 ($p < .05$), and 5.54 ($p < .05$); for Contrast 2 (Ladders vs. Ladders + PALS), it was 1.20, 1.90, and .62 (all *ns*), respectively. Eta squared for the LA, AA, and HA students, respectively, was .14, .18, and .07. For Ladders versus control, Cohen's d for segmentation and blending, respectively, was .71 and 1.57 for LA students, .76 and .47 for AA students, and .31 and .98 for HA students. For Ladders + PALS versus control, Cohen's d for segmentation and blending, respectively, was .46 and .73 for LA students, 1.48 and 1.61 for AA students, and .77 and .71 for HA students. For Ladders + PALS versus Ladders, Cohen's d for segmentation and blending, respectively, was $-.05$ and $-.72$ for LA students, .50 and .80 for AA students, and .44 and .13 for HA students.

On alphabetics (RLS, Word Attack, Word ID, and Spelling), for LA, AA, and HA students, respectively, the $F(4, 27)$ for Contrast 1 (Ladders + PALS vs. Ladders and control) was 2.60 ($p = .06$), 1.60 (*ns*), and .98 (*ns*); for Contrast 2 (Ladders vs. control), it was 1.05, 1.78, and .24 (all *ns*). Eta squared for the LA, AA, and HA students, respectively, was .21, .20, and .08. For Ladders versus control, Cohen's d for the univariate variables ranged from

HA						Across student type					
Ladders + PALS		Ladders		Control		Ladders + PALS		Ladders		Control	
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
6.11	6.02	5.42	4.49	5.03	4.48	3.95	3.79	3.34	2.38	3.02	2.03
22.35	6.73	23.35	6.93	14.16	8.56	17.93	5.12	16.94	6.03	10.18	6.71
12.19	4.32	12.02	4.54	6.50	5.07	9.85	3.53	8.78	3.50	4.17	3.12
33.83	8.64	35.85	11.08	33.22	6.63	18.76	6.62	19.94	7.41	17.88	5.95
10.52	7.20	12.65	8.27	10.17	11.51	6.78	5.03	6.28	6.27	5.26	3.88
2.45	3.91	1.82	3.30	1.34	2.51	0.94	1.61	0.89	1.34	0.54	0.94
5.99	9.06	5.55	7.19	3.72	5.27	2.21	3.09	2.38	2.83	1.45	2.04
33.31	9.72	27.20	9.23	24.65	7.43	25.49	8.22	18.01	8.39	15.81	6.20
9.05	5.41	7.36	6.49	4.29	5.16	4.98	3.06	3.67	3.46	2.03	2.54
18.57	15.03	16.25	12.24	12.54	10.01	9.20	7.36	8.05	6.71	5.47	4.76
55.70	23.74	51.49	29.65	49.77	23.43	40.33	16.97	34.16	18.06	31.03	14.85

RLS = Rapid Letter Sound; WAT = Word Attack; Word ID = Word Identification.

-.09 to .66 for LA students, .15 to .88 for AA students, and -.04 to .45 for HA students. For Ladders + PALS versus control, Cohen's *d* ranged from .23 to .81 for LA students, .81 to 1.38 for AA students, and .73 to 1.01 for HA students. For Ladders + PALS versus Ladders, Cohen's *d* ranged from .11 to .62 for LA students, .34 to .65 for AA students, and .42 to .70 for HA students.

Teacher Satisfaction With PALS

We used a posttreatment questionnaire to ask teachers to indicate their satisfaction with PALS by responding to a 5-point Likert-type scale ranging from 1 (*not at all important*) to 5 (*very important*). The mean teacher response to the question "How important was the What Sound activity to students' reading-readiness achievement?" was 4.55 (*SD* = .82), 4.82 (*SD* = .40), 4.64 (*SD* = .81), and 4.73 (*SD* = .47) for LA, AA, and HA students, and across the three student types, respectively. Identically structured questions were asked of sound boxes, sight words, and reading sentences activities. Corresponding mean responses for sound boxes were 3.73 (*SD* = 1.49), 4.45 (*SD* = .82), 4.73 (*SD* = .65), and 4.36 (*SD* = .92); for sight words, they were 3.91 (*SD* = 1.39), 4.45 (*SD* = .82), 4.64 (*SD* = .67), and 4.64 (*SD* = .67); and for reading sentences, they were 3.40 (*SD* = 1.65), 4.40 (*SD* = .84), 4.55 (*SD* = .69), and 4.50 (*SD* = .85). The mean responses to the question "How important was PALS in helping increase students' overall reading readiness?" were 4.55 (*SD* = .93), 4.64 (*SD* = .50), 4.64 (*SD* = .67), and 4.64 (*SD* = .50). Averaged mean responses to the question "How important was PALS in improving students' social skills?" were 3.90 (*SD* = .99), 3.90 (*SD* = 1.10), 3.80 (*SD* = 1.03), and 4.00 (*SD* = 1.05).

Discussion

Kindergarten Performance

Ladders teachers and Ladders + PALS teachers made regular use of phonological awareness activities; most control teachers

made more sporadic use of them. On posttreatment phonological awareness (segmentation and blending) tasks, Ladders and Ladders + PALS students performed comparably, whereas both treatment groups reliably outperformed controls. Across segmentation and blending tasks, and across LA, AA, and HA students, effect sizes for the Ladders versus control comparison ranged from .46 to 1.30. For the Ladders + PALS versus control contrast, effect sizes ranged from .45 to 2.10. With the exception of one control teacher, only Ladders + PALS teachers provided regular and systematic instruction on beginning decoding skills and gave their students repeated opportunity (by using PALS) to practice these skills. Ladders + PALS students outperformed Ladders and control students on posttreatment alphabetic (RLS, Word ID, Word Attack, and Spelling) tasks. (For HA students, Ladders + PALS > Ladders and controls, $p = .07$.) Across tasks and LA, AA, and HA students, effect sizes for the Ladders + PALS versus control comparison ranged from .08 to 1.42; Ladders + PALS versus Ladders produced effect sizes ranging from .02 to 1.96. Ladders and control students' performance was similar on the alphabetic measures. (These effect sizes appear to be similar to those reported by Bus and van IJzendoorn [1999, Table 2, p. 410]. However, because they do not provide effect sizes for "hybrid" programs such as Ladders + PALS, their findings are not directly comparable to ours.)

This distinct pattern of findings—Ladders and Ladders + PALS students performing better than controls on phonological awareness tasks, and Ladders + PALS students performing best on alphabetic measures—held for LA, AA, and HA students in Title I and non-Title I schools. The specificity of these treatment effects argues persuasively that kindergarten children can indeed be taught phonological awareness by their classroom teachers and that combining phonological awareness training with decoding instruction and practice strengthens beginning reading performance more so than phonological awareness training alone. Fur-

Table 5
Follow-Up Reading Performance by Student Type and Study Group

Compound variable, trial, and measure	LA						AA					
	Ladders + PALS		Ladders		Control		Ladders + PALS		Ladders		Control	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Phonological awareness												
Pretreatment												
Segmentation	1.13	1.03	1.63	2.05	1.80	2.09	4.55	5.34	2.89	2.56	2.43	2.91
Posttreatment												
Segmentation	10.74	7.87	12.73	6.72	7.89	7.57	18.38	5.22	16.32	6.45	8.55	5.53
Blending	8.55	3.57	6.74	4.11	2.85	3.71	9.32	3.51	8.18	4.05	3.77	3.93
Follow-up												
Segmentation	20.62	11.47	21.06	6.12	16.38	7.11	27.87	5.85	24.25	8.49	18.42	6.92
Blending	11.87	3.68	14.14	2.66	9.22	3.60	17.80	3.26	14.47	5.06	12.47	3.37
Alphabetics												
Pretreatment												
RLN												
RLS	2.21	1.97	1.13	1.74	1.72	2.20	7.63	6.36	4.75	5.22	3.79	3.68
WAT	0.04	0.15	0.00	0.00	0.05	0.15	0.48	1.50	0.15	0.26	0.27	0.61
Word ID	0.02	0.08	0.07	0.12	0.27	0.60	0.86	1.07	0.39	0.63	0.63	1.15
Posttreatment												
RLS	18.31	11.15	9.51	7.57	8.75	6.15	25.95	8.01	18.62	10.96	15.05	9.63
WAT	1.43	1.83	0.65	1.23	0.61	1.35	4.46	3.28	2.48	3.45	1.50	2.50
Word ID	2.06	3.33	1.37	1.58	1.29	1.96	7.96	7.47	4.79	5.48	3.21	4.90
WIAT	25.68	10.35	20.10	10.02	20.70	13.84	45.55	22.69	33.64	19.23	28.83	12.39
Follow-up												
RLS	25.25	8.62	20.66	6.11	18.81	7.26	33.51	8.21	30.14	9.87	22.33	7.91
WAT	3.70	3.19	3.41	2.32	1.94	2.14	7.32	5.49	5.44	5.13	3.03	2.14
Word ID	11.35	5.18	10.00	6.00	8.21	5.66	21.01	10.01	15.72	8.09	13.95	7.45
Spelling	58.16	18.71	52.53	17.10	54.05	17.18	79.08	25.33	63.05	24.22	59.95	16.57

Note. LA = low achievers; AA = average achievers; HA = high achievers; PALS = Peer-Assisted Learning Strategies; RLN = Rapid Letter Naming;

ther, the Ladders + PALS children's stronger performance signifies more than a statistically significant difference; as indicated by related effect sizes and NCEs, it represents an educationally important advantage. Such findings are in line with Hatcher et al.'s (1994) conclusion that, "working on phonological skills in isolation is not an optimal method for improving literacy skills" (p. 54).

A related issue, one that is often discussed in the literature (e.g., Bradley & Bryant, 1985; Ehri, 1979), is whether a causal arrow should be drawn from phonological awareness to reading, from reading to phonological awareness, or in both directions. The Ladders children in our study, as mentioned earlier, did not reliably distinguish themselves from controls with respect to their posttreatment performance on alphabetic tasks. However, related effect sizes generally favored the Ladders group (.14 to .33 for LA students, -.45 to .58 for AA students, and .06 to .80 for HA students). Thus, there is a basis (albeit an unsteady one) for believing that phonological awareness can promote early reading. Because of an absence in our study design of a reading-only group, we cannot provide evidence that bears directly on whether reading development spurs phonological awareness. Still, the effect sizes for the Ladders + PALS vs. Ladders comparison were .64 and .69 for AA students' segmentation and blending scores, respectively. This suggests that, for at least the AA students, reading development may positively affect growth in phonological awareness beyond the influence exerted by phonological awareness training. In short, we offer inconsistent evidence that phonological awareness strengthens reading and indirect and incomplete evi-

dence of the reverse; the two in combination suggest, however tentatively, a bidirectional phonological-awareness-early-reading relationship.

What is the contribution of this study, apart from its general affirmation of findings from prior research? Our response is in four parts. First, as mentioned earlier, this investigation is one of a small set of training studies involving preschoolers or kindergartners in which teachers were responsible for student training and treatment implementation. Such studies are vitally important for researchers to determine a treatment's feasibility. By exploring whether teachers can conduct a given treatment easily, enjoyably, accurately, and independently, as well as whether it strengthens student performance, researchers can address the question "Will it work in classrooms?" Research conducted outside classrooms, or inside classrooms with researchers or confederates implementing the treatment, cannot answer this bottom-line question.

Evidence of PALS feasibility in this study comes mostly from the fidelity data. Teachers' averaged accuracy of PALS implementation at Time 1 (Week 9) was 85%; at Time 2 (Week 16), it was 82%. Students' averaged fidelity scores at Times 1 and 2 were 87% and 77%, respectively. This indicates that teachers and students conducted the treatment accurately and frequently. One might ask, "Why frequently?" After all, beyond what they wrote on their calendars, we had no proof of the number of times teachers implemented PALS. Nevertheless, the fidelity data at Time 2 are suggestive: If teachers' implementation was infrequent, it would have been virtually impossible for them and their students to have

HA						Across student type					
Ladders + PALS		Ladders		Control		Ladders + PALS		Ladders		Control	
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
6.10	5.59	5.28	4.51	4.90	5.13	3.93	3.75	3.27	2.26	3.05	2.45
24.04	5.93	22.37	7.81	13.96	8.60	17.72	5.05	17.14	6.07	10.13	6.58
12.48	4.76	11.66	4.91	6.31	5.13	10.55	2.43	8.86	3.59	4.31	3.54
29.88	8.90	26.02	8.82	23.47	7.70	26.12	6.88	23.78	6.32	19.42	4.55
17.33	3.83	16.78	4.33	14.21	4.95	15.70	2.45	15.13	3.29	11.96	3.18
11.32	8.52	12.12	11.87	9.68	6.86	7.05	5.28	6.00	6.13	5.06	4.89
2.64	3.99	1.57	3.25	1.55	2.87	1.06	1.76	0.57	1.13	0.62	1.13
7.09	10.45	5.17	7.13	3.72	6.31	2.66	3.61	1.88	2.41	1.54	2.61
34.71	10.19	25.19	10.39	24.39	8.70	26.33	8.78	18.03	9.15	16.07	6.88
9.59	6.46	6.60	6.78	4.80	6.64	5.16	3.17	3.25	3.56	2.30	3.24
20.66	15.83	14.89	12.67	13.78	12.88	10.22	7.80	7.01	6.20	6.09	6.03
56.68	25.04	51.47	29.58	51.14	31.25	44.55	15.35	35.07	18.25	32.89	17.51
37.34	6.70	32.97	10.05	34.24	7.81	32.04	7.49	27.92	6.39	25.13	6.12
12.69	8.21	8.87	7.61	8.47	9.14	7.90	4.68	5.91	4.70	4.48	4.31
35.23	16.01	28.02	13.33	28.04	13.54	22.53	9.18	17.91	8.61	16.73	6.76
91.98	21.88	81.45	30.45	85.17	28.20	78.39	15.77	65.67	20.52	66.39	14.89

RLS = Rapid Letter Sound; WAT = Word Attack; Word ID = Word Identification.

sustained the high degree of fidelity demonstrated at Time 1. Thus, on the basis of evidence and inference, we conclude that participants found PALS easy to implement, a conclusion in keeping with earlier and more clear-cut findings on the feasibility of implementing PALS in Grades 2 to 5 (e.g., Simmons et al., 1994, Table 4, p. 210).

A second noteworthy feature of this study is the peer-mediated nature of PALS. Absent a necessary component analysis of PALS, we believe peer mediation was the critical ingredient making the Ladders + PALS treatment practical and effective. It promoted practicality because when students learned what to do, they were largely responsible for implementation. Of course, teachers introduced new letter sounds and sight words before PALS sessions, and they monitored the tutoring, encouraging cooperation and providing support when both tutor and tutee were stuck. However, teachers accomplished most of this "off stage." Many teachers said PALS provided them with rare opportunities to observe and reflect on students' behavior and reading performance.

Moreover, PALS's structured, reciprocal, one-to-one interaction between partners (a) permitted frequent opportunity to respond, (b) facilitated immediate corrective feedback, (c) increased academic engaged time, and (d) offered social support and encouragement—features that comply with generally accepted principles of effective instruction. A more constructivist spin might suggest that PALS activities represent the less skilled student's zone of proximal development (Vygotsky, 1978), that is, what the student can do only with the assistance of the more skilled peer. This more

accomplished partner scaffolds the activity by completing a task that the less skilled student could not otherwise do by providing an enabling bit of information, restructuring the task into smaller steps, modeling a correct response, or offering an effective means of producing a response.

In any case, the peer-mediated nature of PALS provided teachers with an organizational strategy for delivering decoding instruction and facilitating practice that was intensive and individualized to a degree that would have been virtually impossible to replicate with teacher-led instruction (see Juel, 1996). We highlight peer mediation because we believe it has potential as a general approach for accelerating student achievement and because it is used infrequently by researchers and practitioners, especially with young children. Because of its infrequent use, there is still much to learn. For example, what specific helping strategies are most useful to kindergarten tutors and tutees? Which reading-related activities are developmentally appropriate for peer mediation and which should be teacher led? Does rate of participation in dyadic versus small-group interaction interact with students' achievement status?

A third way in which this study contributes to the literature is by providing a relatively strong causal claim for the teacher-implemented Ladders and Ladders + PALS treatments. We randomly assigned teachers within schools to the three study groups and documented no pretreatment differences among students or teachers. In our formal statistical comparisons, teacher, not student, was the unit of analysis. With observational data collected in

classrooms, we documented that Ladders teachers and Ladders + PALS teachers (and students) implemented the treatments with fidelity. Treatment children generally did well on outcome measures that reflected the nature and aims of the treatments in which they participated. This specificity of treatment effects was also addressed by using two treatment groups rather than one. A statistically significant difference favoring the Ladders + PALS group over the Ladders group on the alphabetic measures, for example, indicates that the former group's superiority may not be explained by experimenter attention or Hawthorne effects. Our classroom observations and teacher-completed questionnaires indicated an absence of formal beginning reading instruction in most study classrooms, which eliminates another competing explanation of observed effects.

Finally, findings appear generalizable because of a relatively large number of study schools ($N = 8$), teachers ($N = 33$), and students ($N = 404$) and an absence of statistically significant interactions involving treatment group, school type (Title I vs. middle class), or student type (HA vs. AA vs. LA). Likewise, treatment effects were essentially the same in a separate analysis of the 25 children with disabilities in our sample (Fuchs et al., 1999).

Follow-Up Performance (and Other Caveats)

There was a modest weakening of treatment effects on phonological awareness and alphabetic measures over a 5-month stretch from late April (posttreatment testing) to early October (follow-up testing). Ladders and Ladders + PALS students continued to achieve statistically significant, stronger scores than did controls on the phonological awareness tasks, although the effect sizes were somewhat smaller. On the alphabetic measures, an opposite pattern of findings emerged: Ladders + PALS students were no longer reliably different from Ladders and control students (for LA students, $p = .06$); however, the effect sizes favoring the Ladders + PALS group in October were similar to those obtained in April.

Ladders versus controls and Ladders + PALS versus controls generated 12 effect sizes for phonological awareness performance at follow-up (2 group comparisons \times 3 student groups \times 2 measures), 9 of which ranged from .71 to 1.57. The authors of seven studies in the Bus and van IJzendoorn (1999) meta-analysis explored long-term effects of phonological awareness training. At follow up ($M = 8$ months, range = 3 to 12 months), the mean effect size was .48. For the alphabetic tasks, we generated a total of 24 effect sizes at follow-up (2 group comparisons \times 3 student groups \times 4 measures), 22 of which ranged from .34 to 1.38. In eight studies in the Bus and van IJzendoorn meta-analysis, researchers addressed the long-term effects of beginning reading instruction. At follow up ($M = 17.6$ months, range = 6 to 29 months), the mean effect size was .16. Thus, on both phonological awareness and alphabetic measures, Ladders and Ladders + PALS students did considerably better at follow-up than did students in the typical study in Bus and van IJzendoorn's meta-analysis. Yet, this comparison unfairly inflates our study findings on at least two counts: First, our 5-month follow-up was much shorter than most of the studies addressing long-term treatment effects in the meta-analysis; second, Ladders + PALS students were taught reading, whereas students in many of the studies in the meta-analysis were not.

There are more admonitions to be considered. First, missing from our study design was a PALS-only group that, had we the resources to include it, might have helped us better understand relations between development in phonological awareness and emergent reading skills. Second, whereas teachers directed the Ladders activities, PALS was largely peer mediated. A more rigorous comparison of phonological training versus phonological training in combination with beginning decoding activities would exert stronger control over the teacher-directed versus peer-mediated aspect of the treatments. Third, our 15 Ladders activities did not constitute an exclusively phonological awareness treatment. As described earlier, several activities were designed to promote connections between phoneme sounds and letters. Overall, however, our Ladders activities infrequently incorporated letters, a fact that did not escape teachers' attention. On our post-treatment questionnaire, no more than 4 of 11 Ladders teachers indicated that any one of the 15 Ladders activities promoted letter-sound correspondence.

Fourth, we failed to determine interrater agreement for observations of the fidelity with which teachers conducted Ladders. Although we established strong interrater agreement among research staff prior to the collection of these data, we did not document interrater agreement during data collection. As a result, we do not know whether the data on Ladders fidelity suffer from observer drift. Fifth, our testers were aware of children's treatment affiliations. Although we took steps to guard against possible experimenter bias, we have no proof that our testers did not bias their scoring.

Finally, we need to qualify our claim for the generalizability of treatment effects. Although there were no statistically significant interactions involving treatment and type of school or type of student, not all Ladders and Ladders + PALS students responded positively to the treatments. Many of the nonresponders were LA students with and without disabilities. For example, among the 9 children with disabilities participating in Ladders + PALS, pre-to-posttreatment growth on Word ID was as great as 16 points; on Word Attack, it was as great as 11 points. However, on the Word ID and Word Attack subtests, respectively, 4 of 9 and 5 of 9 children made no gains whatsoever (Fuchs et al., 1999). Therefore, Ladders + PALS appears to enhance the capacity of teachers to accommodate many, but not all, children.

Bridging Research and Practice

At various points, we have mentioned the small number of phonological awareness training studies that rely on teachers to conduct the treatment, and we have discussed how this represents an important obstacle in bridging research and practice. The following question remains: Why are there so few studies? We offer three explanations. The first is that numerous educational researchers have rejected the very epistemological basis of the field-based experiment. So-called antipositivists are conducting action research, ethnomethodological research, and other nontraditional types of scholarship. Many traditional researchers have greater interest in theory building than in knowledge application; they require the highly controlled environment of laboratories and have little need for the sometimes helter-skelter ambience of schools and classrooms.

Second, it is costly to involve teachers as treatment implementers. They need materials, training, and on-site technical assistance, which is labor intensive. So is the collection of fidelity-of-treatment data. These and other research-related activities cannot be accomplished without research assistants (often graduate students) who require tuition waivers, monthly stipends, or both. Too few investigators have access to the amount of money necessary to mount large-scale training studies in which teachers are responsible for treatment implementation. This is because of weak federal support of educational R & D, which in Fiscal Year 2000, accounted for less than three tenths of 1% of all federal research expenditures (Fossum et al., 2000). And, whereas such an apportionment no doubt strikes many an American researcher as stingy, the actual dollar amount represents a king's fortune to researchers in the Commonwealth nations and elsewhere.

Third, many researchers we know are reluctant to ask teachers to undergo extensive training, to tolerate the presence of research assistants in their classrooms, to permit evaluations of their treatment implementation, to agree to the random assignment of their students to one of several study groups, and so on. These researchers, or their assistants, often prefer to implement the treatments in the teachers' classes rather than risk losing the teachers and their students. Such caution is understandable, especially given that many districts restrict researchers' access to teachers and students because research is viewed as interfering with the districts' educational mission.

We have been fortunate to have had enough money to conduct training studies in which teachers implement treatments in their classrooms. Nonetheless, we have had to work hard to earn the teachers' respect and cooperation. We have both succeeded and failed at this over the years. Where we have succeeded, we have made teachers our research partners. Before project start up, for example, we explain to them the theoretical and practical purposes of the research and how the study design reflects these purposes. At study's end, we hold meetings in which we encourage them to describe their experiences with, and appraisals of, the treatments. During the summer, we share study findings with them and, with their input, we conceptualize the next steps in the research program (Fuchs & Fuchs, 1998).

We also provide participating teachers with expert and dependable technical assistance by selecting capable and motivated graduate students as research assistants, training them well, and holding them to high standards. Finally, we regularly pitch stories about the study schools to local radio and TV stations and newspapers, helping our school-based partners get the recognition they deserve. Such relationship building is time consuming, is not always easy, and may go beyond some researchers' definitions of their roles. However, by forging personal connections with practitioners, we gain their trust. This in turn permits us to ask them to conform to the dictates of experimental research. Many do so enthusiastically.

There are countless obstacles to discourage researchers from conducting training studies with the highest scientific rigor, just as there are numerous causes for the gap that frequently separates the worlds of research and practice. Nevertheless, most researchers want very much to bridge this gap to make a difference for students and teachers. To them we offer an observation and a belief that guide our own imperfect work: Researchers cannot bridge research and practice by themselves; they need help from admin-

istrators, school boards, teacher unions, parent groups, school faculties, and others. But by conducting first-rate science in schools, it is the researcher who lays the foundation for such bridge work. Conducting first-rate, school-based science is a noble goal worth striving for.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Ball, E. W., & Blachman, B. A. (1991). Does phoneme segmentation training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly*, 26, 49–66.
- Blachman, B. A., Ball, E. W., Black, R. S., & Tangel, D. M. (1994). Kindergarten teachers develop phoneme awareness in low-income, inner-city classrooms: Does it make a difference? *Reading and Writing*, 6, 1–18.
- Bradley, A. (1998). Ed. schools getting heat on reading. *Education Week*, 17(23), 1, 16–17.
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature*, 301, 419–421.
- Bradley, L., & Bryant, P. E. (1985). *Rhyme and reason in reading and spelling*. (International Academy for Research in Learning Disabilities Monograph Series No. 1). Ann Arbor: University of Michigan Press.
- Brady, S., Fowler, A., Stone, B., & Winbury, N. (1994). Training phonological awareness: A study with inner-city kindergarten children. *Annals of Dyslexia*, 44, 26–59.
- Brennan, F., & Ireson, J. (1997). Training phonological awareness: A study to evaluate the effects of a program of metalinguistic games in kindergarten. *Reading and Writing: An Interdisciplinary Journal*, 9, 241–263.
- Bryant, P., & Bradley, L. (1985). *Children's reading problems*. Oxford, England: Blackwell.
- Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology*, 26, 429–438.
- Bus, A. G. (1986). Preparatory reading instruction in kindergarten: Some comparative research into methods of auditory and auditory visual training of phonemic analysis and blending. *Perceptual and Motor Skills*, 62, 11–24.
- Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91, 403–414.
- Byrne, B., & Fielding-Barnsley, R. F. (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.
- Camine, D., Silbert, J., & Kameenui, E. J. (1990). *Direct instruction reading* (2nd ed.). Columbus, OH: Charles E. Merrill.
- Cassidy, J., & Wenrich, J. K. (1998, February/March). What's hot, what's not for 1998. *Reading Today*, 1, 28.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Content, A., Kolinsky, R., Morais, J., & Bertelson, P. (1986). Phonetic segmentation in prereaders: Effect of corrective information. *Journal of Experimental Child Psychology*, 42, 49–72.
- Cunningham, A. E. (1990). Explicit versus implicit instruction in phonemic awareness. *Journal of Experimental Child Psychology*, 50, 429–444.
- Ehri, L. C. (1979). Linguistic insight: Threshold of reading acquisition. In G. Waller & G. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 1, pp. 63–111). New York: Academic Press.
- Elkonin, D. B. (1973). USSR. In J. Downing (Ed.), *Comparative reading:*

- Cross-national studies of behavior and processes in reading and writing* (pp. 551–579). New York: Macmillan.
- Farr, R. C., & Strickland, D. S. (1995). *Treasury of literature: First street*. Orlando, FL: Harcourt Brace.
- Felton, R. H., & Wood, F. B. (1992). A reading level match study of nonword reading skills in poor readers with varying IQ. *Journal of Learning Disabilities, 25*, 318–326.
- Fossum, D., Painter, L. S., Williams, V., Yezzil, A., Newton, E., & Trinkle, D. (2000). *Discovery and innovation*. Santa Monica, CA: RAND.
- Fox, B., & Routh, D. K. (1984). Phonemic analysis and synthesis as word attack skills: Revisited. *Journal of Educational Psychology, 76*, 1059–1064.
- Fuchs, D., & Fuchs, L. S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research, 56*, 243–262.
- Fuchs, D., & Fuchs, L. S. (1998). Researchers and teachers working together to adapt instruction for diverse learners. *Learning Disabilities Research and Practice, 13*, 126–137.
- Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-Assisted Learning Strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34*, 174–206.
- Fuchs, D., Fuchs, L. S., Power, M. H., & Dailey, A. M. (1985). Bias in the assessment of handicapped children. *American Educational Research Journal, 22*, 185–198.
- Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., & Braun, M. (1999, April). *Strengthening kindergartners' reading readiness in Title I and non-Title I schools: The value-added of peer-mediated learning*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Goldman, R., & Lynch, M. E. (1986). *High Hat early reading program*. Circle Pines, MN: American Guidance Service.
- Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. London: Erlbaum.
- Haddock, M. (1976). Effects of an auditory and an auditory-visual method of blending instruction on the ability of prereaders to decode synthetic words. *Journal of Educational Psychology, 68*, 825–831.
- Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development, 65*, 41–57.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- International Reading Association. (1999, June/July). *In search of excellence: Reading Excellence Act moves forward*. Author.
- 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.
- Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly, 31*, 268–289.
- Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*, 243–255.
- Kaestle, C. F. (1993). The awful reputation of education research. *Educational Researcher, 22*, 23–31.
- Kantowitz, B., & Underwood, A. (1999, November). Dyslexia and the new science of reading. *Newsweek, 133*(21), 72–77.
- Levy, B. A., & Lysunchuk, L. (1997). Beginning word recognition: Benefits of training by segmentation and whole word methods. *Scientific Studies of Reading, 1*, 359–387.
- Levy, B. A., & Stewart, L. (1991, April). *Early diagnosis and treatment of reading problems*. Paper presented at the annual meeting of the Society for Research in Child Development, Seattle, WA.
- Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23*, 263–284.
- Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159–173.
- Lyon, G. R. (n.d.). *The NICHD research program in reading development, reading disorders, and reading instruction: A summary of research findings. Keys to successful learning: A national summit on research in learning disabilities*. New York: National Center for Learning Disabilities.
- Mathes, P. G., Howard, J. K., Allen, S., & Fuchs, D. (1998). Peer-Assisted Learning Strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly, 33*, 62–94.
- O'Connor, R. E., Jenkins, J. R., & Slocum, T. A. (1995). Transfer among phonological tasks in kindergarten: Essential instructional content. *Journal of Educational Psychology, 87*, 202–217.
- O'Connor, R. E., Notari-Syverson, N., & Vadasy, P. (1996). Ladders to literacy: The effects of teacher-led phonological activities for kindergarten children with and without disabilities. *Exceptional Children, 63*, 117–130.
- O'Connor, R. E., Notari-Syverson, N., & Vadasy, P. (1998a). First grade effects of teacher-led phonological activities in kindergarten for children with mild disabilities: A follow-up study. *Learning Disabilities Research and Practice, 13*, 43–52.
- O'Connor, R. E., Notari-Syverson, N., & Vadasy, P. (1998b). *Ladders to literacy: A kindergarten activity book*. Baltimore: Paul H. Brookes.
- Olofsson, A., & Lundberg, I. (1983). Can phonemic awareness be trained in kindergarten? *Scandinavian Journal of Psychology, 24*, 35–44.
- Olofsson, A., & Lundberg, I. (1985). Evaluation of long-term effects of phonemic awareness training in kindergarten: Illustration of some methodological problems in evaluation research. *Scandinavian Journal of Psychology, 26*, 21–34.
- Rosenthal, R. (1980). *Experimenter effects in behavioral research* (enlarged paperback edition). New York: Irvington.
- Rosner, J. (1974). Auditory analysis training with prereaders. *The Reading Teacher, 27*, 379–385.
- Schneider, W., Kuspert, P., Roth, E., & Vise, M. (1997). Short- and long-term effects of training phonological awareness in kindergarten: Evidence from two German studies. *Journal of Experimental Child Psychology, 66*, 311–340.
- Share, D., Jorm, A., MacLean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76*, 1309–1324.
- Simmons, D. C., Fuchs, D., Fuchs, L. S., Hodge, J. P., & Mathes, P. G. (1994). Importance of instructional complexity and role reciprocity to classwide peer tutoring. *Learning Disabilities Research and Practice, 9*, 203–212.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stanovich, K. E. (1986). Matthew effect in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–406.
- Stanovich, K. E., Cunningham, A., & Cramer, B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38*, 175–190.
- Stevens, J. (1986). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum.
- Tangel, D. M., & Blachman, B. A. (1992). Effect of phoneme awareness instruction on kindergarten children's invented spelling. *Journal of Reading Behavior, 24*, 233–258.
- The Consortium on Renewing Education. (1998, November). *20/20 vision: A strategy for doubling Americans' academic achievement by the Year*

2020. Nashville, TN: Vanderbilt University, Peabody Center for Education Policy.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading, 1*, 217–234.
- Torneus, M. (1984). Phonological awareness and reading: A chicken and egg problem? *Journal of Educational Psychology, 76*, 1346–1358.
- Troia, G. A. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. *Reading Research Quarterly, 34*, 28–52.
- Tunmer, W. E., & Hoover, W. A. (1992). Cognitive and linguistic factors in learning to read. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 175–214). Hillsdale, NJ: Erlbaum.
- Tur-Kaspa, H., Mioduser, D., & Leitner, I. (in press). The effect of early reading intervention on phonological awareness, word recognition, and letter-naming skills of kindergarten children at risk for reading failure. *Journal of Special Education*.
- Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33*, 321–363.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192–212.
- Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85*, 83–103.
- Wallach, M. A., & Wallach, L. (1976). *Teaching all children to read*. Chicago: University of Chicago Press.
- Weiner, S. (1994). Effects of phonemic awareness training on low- and middle-achieving first graders' phonemic awareness and reading ability. *Journal of Reading Behavior, 26*, 277–300.
- Whitehurst, G. J., Epstein, J. N., Angell, A. L., Payne, A. C., Crone, D. A., & Fischel, J. E. (1994). Outcomes of an emergent literacy intervention in Head Start. *Journal of Educational Psychology, 86*, 542–555.
- Williams, J. P. (1980). Teaching decoding with an emphasis on phoneme analysis and phoneme blending. *Journal of Educational Psychology, 72*, 1–15.
- Wolf, M. (1991). Naming speed and reading: The contribution of the cognitive neurosciences. *Reading Research Quarterly, 26*, 123–141.
- Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests—Revised*. Circle Pines, MN: American Guidance Service.
- Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23*, 159–177.

Received May 19, 2000

Revision received September 28, 2000

Accepted September 28, 2000 ■

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.