Computer-Based Remedial Training in Phoneme Awareness and Phonological Decoding: Effects on the Posttraining Development....

Article *in* Scientific Studies of Reading · July 1997

DOI: 10.1207/s1532799xssr0103_4

4 authors, including:


Richard K Olson
University of Colorado Boulder


Barbara Wise
Linguistic Remedies for Reading Disabilities

SCIENTIFIC STUDIES OF READING, *1*(3), 235–253

Computer-Based Remedial Training in Phoneme Awareness and Phonological Decoding: Effects on the Posttraining Development of Word Recognition

Richard K. Olson, Barbara Wise,
Jerry Ring, and Mina Johnson
University of Colorado, Boulder

This article reports results from 1- and 2-year follow-up tests for children with reading problems who had previously received 25 hr of training in 1 of 2 computer-based remediation programs. Both programs included accurate speech and decoding support during story reading, but different supplementary training, either in explicit phonological processes or in comprehension strategies. Children initially trained in phonological skills made greater gains in phonological awareness, phonological decoding, and untimed word recognition at the end of training. The other children, who had spent more time reading stories with speech feedback for difficult words, gained more in time-limited word recognition. One year later with no further training, the explicit phonological group remained stronger in phonological decoding and phoneme awareness, but its level of word recognition was no greater than in the other group. There were no significant group differences after 2 years. Both of the computer-based training programs had yielded rapid growth in word recognition standard scores during 50 half-hour training sessions over 4 months, although much more training is needed to bring many children with reading disability into the normal range.

In this article, we report new results from 1- and 2-year follow-up tests for a computer-based reading remediation study that was conducted during the spring semester of 1994 (Wise & Olson, 1995). Both training groups in this study received phonological decoding and speech support to help them read difficult words in stories on computers, but only one group received explicit training in phonological

Requests for reprints should be sent to Richard Olson, Department of Psychology, Box 345, University of Colorado, Boulder, CO 80309–0345. E-mail: rolson@psych.colorado.edu

analysis (the PA group). Supplementary training in explicit phonological analysis was compared to a training condition in comprehension strategies (the CS group), which allowed for more time spent reading stories on the computer with speech and decoding support for difficult words.

Both of the training conditions supported strong standard-score gains of 7.5 to 10 points on two tests of word recognition, after only 25 hr of training distributed over 4 months. The PA group attained significantly greater gains in phoneme awareness, phonological decoding (nonword reading), and untimed word recognition. However, the CS group, which had spent more of its training time reading stories on the computer with speech and decoding support, showed significantly greater gains in time-limited word recognition. Wise and Olson (1995) hypothesized that the PA group's superior phonological skills at the end of training would be maintained after training and might lead to more rapid growth in word recognition at 1- and 2-year follow-up tests.

The new follow-up data for the Wise and Olson (1995) study, reported in this article, shows that the PA group did maintain a significant advantage in phoneme awareness and phonological decoding at least 1 year after training ended. However, the PA group's continued superiority in these skills did not result in more rapid gains in word recognition after training, without further support or instruction in its application. We suggest ways for increasing transfer from improved phonological skills to corresponding improvement in word recognition after training. We also highlight the value of using computer-based training in research on methods and in providing efficient and individualized support for children with reading disabilities.

Our interest in phoneme awareness (the explicit awareness of and ability to manipulate phonemes within spoken syllables) and phonological decoding (measured by nonword reading) comes from correlational and behavioral–genetic evidence linking deficits in these skills to reading disability (cf. Olson, Wise, Conners, Rack, & Fulker, 1989; Rack, Snowling, & Olson, 1992). Jorm and Share (1983) and Share (1995) articulated why good phonological decoding is important for normal reading development. They pointed out that phonological decoding can provide a self-teaching mechanism that can lead to and support accurate recognition of unfamiliar printed words. The mechanism is particularly important when children encounter unfamiliar words in independent reading. Although many English words are not completely regular in their spelling, Share argued that they are regular enough that partial letter–sound knowledge can be combined with contextual information and oral–lexical knowledge to produce correct decoding and recognition. Share said that after a few such correct decodings of a new word, that word becomes more automatically accessible through the normal reader's sight vocabulary, without the continued need for explicit phonological decoding

Children with poor phoneme awareness and phonological decoding skills are likely to make many errors when attempting to decode new words on their own. If these errors are not corrected, they result in negative learning trials that constrain the development of a sight vocabulary for those words. Uncorrected decoding errors also reinforce inappropriate print–sound associations that hinder the inductive development of better phonological decoding skills during reading (Jorm & Share, 1983). Therefore, improving phonological decoding is a major goal of many instructional programs.

Our initial computer-based remediation studies were motivated by observations of twins with reading disabilities who were participating in a behavioral–genetic study (Olson et al., 1989). The children made frequent decoding errors when reading stories at their age–grade level. In addition to problems created by these negative learning trials, the children's obvious frustration in reading probably led to much less reading practice as compared to normal readers. Thus, the significant genetic influence on many children's deficits in phonological decoding and word recognition may have been mediated in part by errors in independent reading and by lack of reading practice. We thought that talking computers could help provide structured and accurate reading practice with decoding support for children with reading problems.

In our first training study with the talking computers, third- to sixth-grade children, who were below the local 10th percentile in word recognition, were pulled from their regular remedial reading or language arts classtime to read on the computer for a half hour, 3 to 4 days per week, during one semester (Olson & Wise, 1992; Wise et al., 1989). The program was titled "Reading with Orthographic and Speech Segmentation" (ROSS).The children in the ROSS program read interesting stories on the computer and were trained to target difficult words with a mouse for decoding assistance so they could correctly identify the words. The difficulty level of the stories was continually adjusted, so that children required assistance on 2% to 5% of the words. Targeted words were immediately highlighted in reverse video and pronounced by the computer using Digital Equipment Corporations' (Marlboro, MA) DECtalk speech synthesizer. Children trained on ROSS for a total of only 14 hr over 3 months, but they showed about twice the gains in word recognition when compared to an untrained control group (.6 vs. .3 grade level).

Although the results from only 14 hr of training on ROSS were encouraging, it was clear that much more practice in accurate reading would be needed to bring these children up to local reading norms. Most relevant to this study, significant correlations between pretest phoneme awareness and gains in phonological decoding and word recognition suggested that explicit training to improve phonological skills might also improve children's gains from the story reading: Participants who started in the upper half of the training-group distribution in phoneme awareness gained about twice as much as those in the lower half of the distribution (Olson & Wise, 1992).

Explicit training in phoneme awareness and phonological decoding was included in a study conducted during the spring semester of the 1993–94 school year (Wise & Olson, 1995). Data from 1- and 2-year follow-up tests, with no further training or individualized support, are presented for the first time in this article. The rationale, methods, and immediate posttraining results of the Wise and Olson study are presented here in enough detail to understand the significance of the follow-up data. Wise and Olson compared two computer-based training programs in four Boulder schools for 103 second- to fifth-grade poor readers who were in the lower 10% of their class in word decoding and scored above 90 on the Wechsler IQ test (Wechsler, 1974). English was their first language. Children were assigned to groups of 3 students from one class to ease scheduling problems. The groups were balanced by participant variables and pretest scores and then were semirandomly assigned to a training program. Each school provided a quiet room to house the trainer, 3 children, and their computers. Scheduling, pretesting, and familiarization with the computer was completed by the end of the fall semester. Training in the programs continued from January to the start of May. High levels of attendance resulted in an average of 50 half-hour training sessions during the semester.

The two training programs both provided phonological decoding and speech support for unknown words in story reading, as in the earlier studies, but differed in their relative emphasis and time spent on explicit subword analysis of speech and print, and on reading stories with computer-speech support for decoding difficulties. The PA condition included many of the training procedures of the Lindamood Auditory Discrimination in Depth (ADD) program (Lindamood & Lindamood, 1975). The early training sessions concentrated on small-group activities to help children discover the articulatory movements for different phonemes and associate the appropriate letter symbols, sounds, mouth pictures, and labels for those articulations. For example, the first session included motivation and introduction of the program's first two "quiet and noisy" pairs: "lip poppers */p,b/* and tip tappers */t,d/*."

These small-group activities were supplemented with practice of the articulatory concepts for consonants and vowels using a computer program, under development by the Lindamood-Bell Company, that used animated mouth pictures. Children also used three computer programs for encoding, decoding, and the manipulation of sounds, which were developed in our laboratory. In one program, participants manipulated letter symbols having constant phoneme representations to reflect changes in syllables spoken by the computer (e.g., "If this says viz, show me vooz"). In the second program, children explored English orthographic patterns and print-to-sound relations in a spelling exploration game. In both these programs, the computer could pronounce both correct and incorrect attempts, and children could modify their attempts accordingly (Wise & Olson, 1992). In a third program, children chose among four nonwords on the computer screen to match a nonword spoken by the computer. (These nonwords did not overlap with nonwords in the pre- and posttraining tests.) In these three programs, the computer automatically increased or decreased the difficulty level of exercises according to the child's performance.

When children in the PA group succeeded with the CVC level in the first computer exercise, they began spending 7 to 15 min of their training time reading stories in the ROSS program. PA participants were encouraged to use their analytic decoding strategies when encountering unknown words in the stories. To help promote the use of these strategies, targeted words were highlighted as follows: The onset and rime units of regular one-syllable words were highlighted in alternating blue and green backgrounds with the first mouse click. Special exception words such as said were not segmented and were highlighted with all letters in one color. Multisyllable words were segmented into syllables. Children were encouraged to "use the colors" to help them sound out each segment, and then click a second time to hear sequential pronunciations of the segments. A third, optional mouse click of a different button pronounced the whole word as a unit. A multiple-choice question occurred at reasonable breaks in the story. A test of 15 of a session's targeted words was presented at the end of each session, with words in their color-segmented format. Training time for the PA group averaged 8.4 hr reading on ROSS, 10 hr on analytic computer exercises, and 7 hr in small-group activities.

The analytic training of the PA condition was contrasted with a condition that included the same trainers, and children from the same schools, but instead trained the comprehension strategies of Palinscar and Brown's (1984) reciprocal-teaching program. As in the PA condition, early training sessions included small-group activities. However, instead of focusing on analytic phonological skills, children in the CS condition read stories in books and on the computer as a group, and practiced comprehension strategies of prediction, question generation, clarification, and summarization. Using reciprocal teaching, the trainer first modeled the strategies, and then the students gradually took over the discussion-leader role. These small-group activities proved to be very popular with the children. In addition, the CS participants spent all of their individualized computer time reading on ROSS: one-third of that time reading aloud with the trainer, and two-thirds doing independent silent reading. The CS Children were also encouraged to target any word they were unsure of and to use the colors to help them sound out the word before clicking the mouse again for computer pronunciation of the segments, and again for the whole-word pronunciation if needed. CS participants' total reading time on ROSS averaged 18 hr, and much of their small-group time off the computer was spent reading and using their cognitive strategies with printed stories.

The primary reason for including the CS training condition was to have a comparison group that spent most of its time reading stories with decoding assistance for unfamiliar words, but that included the same amount of time with interesting small-group activities. Both small-group and individualized computer time was the same for the two groups. What was different, and what we wanted to compare, were the effects of time spent in explicit phonological instruction and practice versus spending that practice time in accurate, phonologically supported story reading.

We controlled for any differences in trainers' general effectiveness in teaching the children by having the same trainers administer both programs. Before entering the schools, the trainers took classes in both methods (from consultants mentioned in the acknowledgments and from the second author) and role-played the PA and CS procedures with each other and with pilot children throughout the fall, with critiques from each other and from the second author. Both methods used a guided discovery, or "teaching-by-questioning" approach, so no difference in general teaching style was required; yet the content of the small groups and the presence of the analytic exercises was quite different. While training children in the schools, adherence to the methods was monitored in weekly lab meetings and in Barbara Wise's visits to each school, initially twice a month, and later once a month. By design, both groups in the study received necessary decoding support in the story reading, but only the PA group received explicit, analytic phonological training.

Significant differences in PA and CS training effects on phoneme awareness, phonological decoding, and word recognition support that the treatment conditions were distinct in their instruction and benefits to phonological skills. Posttests at the end of training revealed that the PA group, when compared to the CS group, demonstrated substantially greater gains in measures of phoneme awareness and phonological decoding (see Table 1). Measures are described in

TABLE 1
Pretest Scores, Gains, and Effect Sizes From Pretest to End of Training

	Pretest		*Gain*				
Measure	*CS*[a]	*PA*[b]	*CS*[a]	*SD*	*PA*[b]	*SD*	*Effect Size*
LAC II (# correct)	5.3	5.3	1.8	3.4	5.2	3.6	1.00*
Phoneme (% correct)	32.8	32.8	6.0	18.0	19.9	19.0	.77*
Nonword (% correct)	24.5	25.3	14.1	9.0	27.0	15.5	1.29*
TL Rec. (# correct)	25.4	24.0	17.1	9.1	12.5	9.1	–.51*
PIAT GE	2.5	2.4	.9	.74	1.1	.64	.27
PIAT SS	85.8	85.9	7.5	6.0	10.0	7.9	.28
WRAT GE	2.7	2.7	1.0	.68	1.1	.73	.15
WRAT SS	74.9	76.0	7.6	8.0	8.8	8.1	.15

Note. CS = comprehension strategies; PA = phonological analysis; LAC II = Lindamood Auditory Conceptualization; TL Rec. = time-limited word recognition; PIAT = Peabody Individual Achievement Test; GE = grade equivalent; SS = standard score; WRAT = Wide Range Achievement Test.

[a]$n = 45$. [b]$n = 58$.

$*p < .01$.

the Method section. The CS group showed a significant advantage on ROSS daily comprehension questions when they were engaged with the trainer in discussions using the strategies, but not when reading independently.

Gains on measures of isolated word recognition showed mixed results. The PA group significantly outperformed the CS group in percentage-correct performance on individualized daily (88% vs. 83%), monthly (74% vs. 62%), and end-of-semester (84% vs. 78%) untimed tests of words they had previously targeted in the ROSS stories. The PA group showed slightly but not significantly ($p > .05$) greater gains in the untimed Peabody Individual Achievement Test (PIAT; Dunn & Markwardt, 1970) and Wide Range Achievement Test (WRAT; Jastak & Jastak, 1978) tests for word recognition. (We now have data from a second study indicating that the PA group's advantage in the WRAT and PIAT are statistically significant, $p < .05$, in a larger sample.) In contrast, the PA group had significantly ($p < .01$) smaller gains in an experimental measure of time-limited (2-sec exposure) word recognition. Wise and Olson (1995) suggested that the PA participants could use their newly acquired phonological skills in the PIAT, WRAT, and individualized ROSS tests because these tests had no time constraints, but their slow and analytic skills put them at a disadvantage in the time-limited test. Overall, we were encouraged to see standard-score gains of 7.5 points for the CS children and 9 to 10 points for the PA children on the WRAT and PIAT tests (see Table 1). Although we do not as yet have local normative data on these tests, our participants' improvement in standard scores on national norms suggested that both training conditions substantially accelerated students' growth in word recognition during the modest 25 hr of training over 4 months. This follow-up study considers whether this improved growth rate and the significant treatment differences in phoneme awareness and phonological decoding continued after training ended.

Wise and Olson (1995) were quite pleased with the PA group's strong gains in phonological skills and in their better performance on the targeted word tests from the ROSS program, but were puzzled that the PA group did not show similarly better performance on the word-recognition posttests. Wise and Olson suggested several possible explanations for their paradoxical results, including the need for (a) more training time to reach local grade norms, (b) practice in learning to vary vowel pronunciation for irregular words, and (c) automaticity training for word recognition and phonological decoding to improve transfer to accurate and rapid word recognition. We will return to these possible explanations in the discussion. A fourth explanation is most relevant to this follow-up study: Wise and Olson suggested that the PA participants may not have had enough practice integrating their improved phonological skills into reading by the end of training, because they had averaged only 8 hr reading stories on ROSS. For this follow-up study, we hypothesized that PA participants would retain their superior phonological skills over the follow-up period, and that after further reading experience during this period (outside of the ROSS program), the PA participants' superior phonological skills might become integrated into the normal reading process to support more rapid growth in word recognition.

METHOD

Participants

The initial Wise and Olson (1995) study included 58 children in the PA condition and 45 children in the CS condition. The groups were well balanced in mean age at pretest (PA = 8.97 years; CS = 8.96 years) and in their pretest scores (see Table 1). All participants were from the second to fifth grades, had Wechsler (1974) full-scale IQ scores above 90 (PA $M = 102$; CS $M = 101$), and were in approximately the lower 10% of their class in word decoding. Some children were above the 10th percentile on the national norms for the WRAT and PIAT tests. However, the local reading percentiles of all third graders in our experimental schools averaged 70% on the nationally standardized California Achievement Test. Thus, most of our participants were below the local 10th percentile.

The WRAT and PIAT tests yielded similar mean grade levels for the groups, but the WRAT standard scores were significantly lower. This puzzling result is a direct consequence of differences in the tests' standardization tables.

At the first follow-up testing in the following spring semester (March, 1995), about 10 months after the end-of-training tests, participants were in the third to sixth grades. Twenty participants from the CS condition and 27 participants from the PA condition were available for the first-year follow-up testing. The sample was reduced primarily by the exclusion of one of the four schools because of its participation in a follow-up training program and by the previous fifth graders in all schools moving on to sixth-grade middle school and being unavailable for testing. Thus, the average pretest ages were lower in the follow-up sample (8.5 years). There were an additional 8 participants from the PA group and 6 participants from the CS group whose parents did not return permission forms. The differences between the pretest scores of these 14 participants and the 47 who participated in the 1st-year follow-up testing were not significant ($p > .05$). However, the average gain scores from the pretests to the end of training did differ for the two groups. The 14 participants who did not participate in follow-up testing had shown significantly higher standard-score gains on the PIAT (14.0 vs. 7.0) and WRAT (15.7 vs. 8.0), and in number correct on our time-limited measure of word recognition (19.2 vs. 13.7; all $p < .05$). There were nonsignificant trends in the same direction for gains on our phoneme awareness and nonword reading measures. Some parents or children, or both, may have declined to participate in follow-up testing because they were no longer concerned about a reading problem. One would expect follow-up gains of these children to be higher, and their maintenance of rate of gain better, than for children who had gained less with training. Many of the high gainers who were not retested at follow-up may have been children who only needed modest environmental support to become normal readers (Vellutino, Scanlon, & Denkla, 1996), or children who responded better to treatment.

At the 2nd-year follow-up testing in March of 1996, 19 PA and 17 CS participants were tested who had been tested in the previous follow-up. Again, it was the children from the excluded school and those who moved on to middle school (now the original fourth and fifth graders) who were disproportionately lost from the sample. This meant that the average pretest age (8.3 years) and grade (second and third) of the 2-year follow-up participants were lower than for the posttraining (Table 1) and 1-year follow-up (Table 2) samples. There were no significant age or initial pretest differences between the groups (see Table 3), and there were no significant gain score differences from participants who were tested in the 1-year but not the 2-year follow-up.

Measures

The following measures from pre- and posttesting were included in both the 1st- and 2nd-year follow-up testing.

Phoneme Deletion. This experimental test included 40 tape-recorded trials. An example trial would say "Say *prot* [participant says *prot*] ... Now say *prot*

TABLE 2
Pretest Scores, Gains, and Effect Sizes at End of Training and 1-Year Follow-Up

	Pretest		*Gain*				
Measure	*CS*[a]	*PA*[b]	*CS*[a]	*SD*	*PA*[b]	*SD*	*Effect Size*
Phoneme (% correct)	24.6	28.6	7.0	13.6	18.4	18.5	.84*
Follow-up			18.4	19.3	24.5	14.7	.32
Nonword (% correct)	21.5	23.7	13.8	10.6	27.4	16.5	1.28**
Follow-up			16.6	13.9	29.5	18.3	.93**
TL Rec. (# correct)	21.1	21.0	15.1	7.2	12.6	10.0	−.35
Follow-up			21.3	10.5	21.2	13.2	.01
PIAT GE	2.4	2.3	.6	.59	1.0	.59	.63
Follow-up			1.1	.78	1.2	.65	.13
PIAT SS	89.4	87.4	4.8	6.5	8.7	7.7	.60
Follow-up			.9	7.5	2.7	7.3	.24

Note. First row for each measure indicates pretest to end-of-training gains; second row is for pretest to 1-year follow-up gains. TL Rec. = time-limited word recognition; PIAT = Peabody Individual Achievement Test; GE = grade equivalent; SS = standard score.
[a]$n = 20$. [b]$n = 27$.
*$p < .05$. **$p < .01$.

TABLE 3
Pretest Scores, Gains, and Effect Sizes at End of Training, 1-Year, and 2-Year Follow-Up

	Pretest		*Gain*				
Measure	*CS*[a]	*PA*[b]	*CS*[a]	*SD*	*PA*[b]	*SD*	*Effect Size*
LAC II (# correct)	4.6	5.2	1.5	3.5	5.8	4.5	1.21**
2-year follow-up			3.4	4.2	4.4	4.5	.24
Phoneme (% correct)	25.5	26.1	4.1	11.1	19.2	20.9	1.36**
1-year follow-up			13.9	14.8	25.6	16.0	.79*
2-year follow-up			23.6	15.5	32.7	13.7	.59
Nonword (% correct)	20.5	25.2	13.8	10.0	31.3	17.3	1.74**
1-year follow-up			16.6	14.6	32.8	19.6	1.11**
2-year follow-up			27.7	17.8	34.5	22.2	.38
TL Rec. (# correct)	20.0	21.0	14.6	6.6	13.4	9.8	–.18
1-year follow-up			20.9	11.4	23.3	10.3	.21
2-year follow-up			35.8	13.2	38.7	16.9	.22
PIAT GE	2.3	2.3	.7	.60	1.0	.50	.48
1-year follow-up			1.1	.84	1.4	.54	.40
2-year follow-up			1.9	.81	2.1	.99	.25
PIAT SS	88.9	88.6	5.5	6.5	9.7	7.3	.64
1-year follow-up			.9	8.1	4.3	7.7	.43
2-year follow-up			.9	7.8	–.1	14.5	–.06
WRAT GE	2.4	2.5	.9	.60	1.1	.50	.34
2-year follow-up			2.5	.85	2.5	1.2	.00
WRAT SS	75.8	77.3	6.5	6.6	9.2	5.4	.41
2-year follow-up			10.9	7.4	9.7	12.3	–.16

Note. First row for each measure indicates pretest to end-of-training gains; second and third rows are for pretest to 1- and 2-year follow-up gains. LAC II = Lindamood Auditory Conceptualization; TL Rec. = Time-limited word recognition; PIAT = Peabody Individual Achievement Test; GE = grade equivalent; SS = standard score; WRAT = Wide Range Achievement Test.

[a]$n = 17$. [b]$n = 19$.

$^{*}p < .05$. $^{**}p < .01$.

without the /r/." The participant should say *pot* within a 6-sec period before the next trial. All trials were nonwords that were to be transformed into words after the deletion. Deletions varied across the initial, medial, and final positions of the nonword and from single consonant or consonant clusters. Yopp (1988) showed that phoneme-deletion tasks are among the most reliable and valid measures of phoneme awareness.

Nonword Reading. Phonological decoding was measured by a 40-item test of pronounceable nonwords (e.g., *niss, kiver*). All items appeared one at a time on the computer screen, with no time limit, and were scored online by the tester.

Time-Limited Word Recognition. This experimental test (TL Rec. in Tables 1 to 3) included a difficulty-ordered list of words presented individually for 2 sec on the computer screen. Testing stopped when participants made five errors on the last seven trials.

PIAT Word Recognition. The 66 words from the PIAT (Dunn & Markwardt, 1970) test for word recognition were presented individually, with no time limit, on the computer. Testing stopped when participants made six errors on the last seven trials. Raw scores were converted to grade-level and standard scores according to published national norms.

Additional measures at pretest and the 2-year follow-up. Two additional measures were included in the Wise and Olson (1995) study and in the 2nd-year but not the 1st-year follow-up test battery (due to constraints on testing time). They were the Lindamood Auditory Conceptualization (LAC) test of phoneme awareness (Lindamood & Lindamood, 1975), and the WRAT test of word recognition (Jastak & Jastak, 1978). The LAC test assessed phoneme awareness by having children order colored blocks to represent the addition, deletion, substitution, and reordering of phonemes in spoken syllables. The number correct out of the 18 trials in Part II of the LAC test are reported here because only that part of the test measures phoneme awareness. The WRAT test of word recognition was given in its usual article format. Raw scores on the WRAT were converted to grade-level and standard scores according to the published national norms.

Procedure

The training procedures for the Wise and Olson (1995) study are described earlier in this article. Trainers were thoroughly instructed to support the PA and CS programs with equal levels of enthusiasm and intensity, and to maintain fidelity of the different treatment conditions. All testing was conducted in a quiet room of the child's school by trained testers. The follow-up measures were given during a 1-hr session or 2 half-hour sessions in March of 1995 and 1996. Test order was the same for all participants.

RESULTS

The minimum alpha level for significance on all tests was set at .05, two-tailed. Exact p values are given for contrasts that approached significance.

One-Year Follow-Up

The mean pretest scores, gains at the end of training, gains at follow-up, and effect sizes are presented in Table 2. Effect sizes are based on subtracting CS gains from PA gains, and dividing the difference by the standard deviation for the CS group's gains. We did not have an untrained control group to use for comparison of treatment effects. Thus, the effect sizes reflect only the relative differences in gains between two strong interventions.

Univariate analyses of variance (ANOVAS) were conducted on each measure, first with initial pretest and end-of-training posttest scores, and next with initial pretest and 1-year follow-up test as repeated measures. As for the full end-of-training sample, this sample showed highly significant ($p < .001$) gains on all measures from pretest to the end of training. The significant gains in PIAT standard scores indicated that in both conditions, the children were rapidly improving their word recognition during training, relative to the national norms.

A significant end-of-training treatment-condition difference in this smaller follow-up sample was found for nonword reading, $F(1, 45) = 10.45, p < .01$, and for phoneme deletion, $F(1, 45) = 5.52, p < .05$. These results are similar to those for the full sample from pretest to end-of-training (see Table 1). However, the full sample's significant CS group advantage in time-limited word recognition was not replicated in this smaller sample, $F(1, 45) = .94, p = .34$. As in the full sample, the PA group in the follow-up sample showed a nearly significant advantage in gains on PIAT grade levels, $F(1, 45) = 3.36, p = .07$, and standard scores, $F(1, 45) = 3.43$, $p = .07$.

Analyses of the initial pretests and 1-year follow-up tests revealed that the PA group maintained a significant superiority in nonword reading, $F(1, 45) = 6.91, p < .05$, but the previously observed advantage in phoneme deletion was no longer significant, $F(1, 45) = 1.45, p = .24$. Despite the PA group's continued superiority in nonword reading over the year after training, it did not exhibit greater growth in word recognition. The effect sizes in Table 2 show that the PA group's nearly significant advantage in gains on the PIAT at the end of training was much smaller and did not approach significance at the 1-year follow-up test. Moreover, there was virtually no group difference in follow-up gains on the time-limited test of word recognition.

Trend analyses were conducted with data from four test times that included pretests, midtraining, end-of-training, and 1-year follow-up tests. (No midtraining test was available for phoneme deletion.) The polynomial trend contrasts were coded for the actual time intervals between tests. The linear main effect of test time was highly significant ($p < .001$) for phoneme deletion, nonword reading, time-limited word recognition, and PIAT grade level, but not for PIAT standard scores ($p = .92$). The linear interaction with training condition was significant only for phonological decoding, $F(1, 45) = 4.95, p < .05$.

It can be seen in Table 2 that there are some quadratic trends that reflect more rapid gains during than after training, most notably in the PA group. The quadratic main effect (average growth of PA and CS conditions combined) was significant for all measures ($p < .01$). In addition, there were several interesting interactions between quadratic effects and training condition. These were significant for PIAT grade level, $F(1, 45) = 4.28, p < .05$; PIAT standard scores, $F(1, 45) = 4.13, p < .05$; nonwords, $F(1, 45) = 4.81, p < .05$; and phoneme deletion, $F(1, 45) = 4.81, p < .05$. For these measures, in the PA condition, growth was more rapid during training and followed by similar or slower growth rates following training, when compared to the CS group. The opposite pattern of quadratic treatment interaction was observed for time-limited word recognition. Growth during training in the CS condition was slightly faster and followed by slightly slower growth after training, when compared to the PA group. However, this trend did not reach significance, $F(1, 45) = 1.56, p = .22$.

Table 2 shows that both the CS and PA groups' PIAT standard scores increased during training and declined after training to standardized levels slightly above their pretest levels. The increase in the CS and PA groups' standard scores during training indicated that the children improved their relative standing in the class and rate of reading development. However, when they were tested a year later without further individual training and support, they had not maintained their accelerated growth on the PIAT. A different pattern of results was found for the WRAT at the 2-year follow-up, where standard-score gains were maintained.

Two-Year Follow-Up

The 2-year follow-up results are presented in Table 3, along with the 1-year follow-up and end-of-training results for this reduced sample. The same analyses were performed as for the previous 1-year follow-up sample, with the LAC II and WRAT measures added to the test battery. The sample size was reduced from the previous 1-year follow-up sample by 11 participants (now, PA $n = 19$, CS $n = 17$). For the univariate ANOVAs, a significant training condition difference was found for (a) nonword reading at the end of training, $F(1, 34) = 13.45, p < .01$; at 1-year follow-up, $F(1, 34) = 7.76, p < .01$; but not at 2-year follow-up, $F(1, 34) = 1.00, p = .32$; (b) phoneme deletion at the end of training, $F(1, 34) = 7.00, p < .05$, and at 1-year follow-up, $F(1, 34) = 5.14, p < .05$; but the difference was not quite significant at 2-year follow-up, $F(1, 34) = 3.47, p = .07$; and (c) the LAC II test at the end of training, $F(1, 34) = 9.69, p < .01$; but not at 2-year follow-up, $F(1, 34) = .44, p = .51$. Training condition effects for all measures of word recognition were nonsignificant at the end of training and at the 1-year and 2-year follow-ups, although the end-of-training advantage for the PA group's PIAT standard scores approached significance, $F(1, 34) = 2.59, p = .12$.

Linear and quadratic trend analyses were based on five test times for the nonword and PIAT tests. The LAC II and WRAT tests were not given at the 1-year follow-up, and the WRAT and phoneme-deletion tests were not given as midtraining tests. The linear and quadratic trends were generally similar to those in the 1-year follow-up. All measures in the 2-year follow-up, except for the PIAT standard scores, had highly significant linear trends but no significant linear interactions with treatment. All quadratic trends were significant, and again there were quadratic interactions with training condition. However, with the reduction in size for the 2-year sample, the quadratic interaction with treatment condition now only approached significance for the PIAT, $F(1, 33) = 3.45$, $p = .07$, and WRAT, $F(1, 34) = .2.42$, $p = .13$, standard scores. The quadratic interactions with treatment were significant for nonwords, $F(1, 34) = 17.14$, $p < .01$; the LAC test of phoneme awareness, $F(1, 34) = 12.57$, $p < .01$; and were almost significant for phoneme deletion, $F(1, 34) = 3.82$, $p = .06$. As in the previous 1-year follow-up sample, these interactions were due to the PA condition's more rapid growth during training, followed by a deceleration of rate of growth and relative class position (shown by a drop in the LAC II raw-score and PIAT standard-score levels), compared to more linear growth during and after training for the CS group.

Although students' absolute skills increased after training, these gains were at much lower rates than during training. There was a decline from the end of training to the 2-year follow-up in the PIAT standard scores for both the PA and CS conditions. Also, there was a greater increase during training followed by a more rapid decrease after training in the PA group, compared to the CS group (significant in the 1-year sample, but not quite significant in the 2-year sample). However, this pattern was not replicated for the WRAT standard scores. That is, there was no decline for the CS or PA groups in their WRAT standard scores following the gains made during training (see Table 3).

Finally, to address the question of stability of individual differences within grade across test times, participants' grade-adjusted pretest scores were correlated with their end-of-training, 1-year, and 2-year follow-up tests within the PA and CS groups. Most correlations were significant and surprisingly high for both groups, in view of the restricted range of the sample and time between tests. For example, for nonwords in the CS and PA groups respectively, pretest to the end of training $r = .74$ and .74, pretest to 1-year follow-up $r = .56$ and .69, and pretest to 2-year follow-up $r = .46$ and .51 (all $p < .01$). The only major exceptions to this pattern were the low and nonsignificant correlations for the LAC II measure.

DISCUSSION

Taken together, the end-of-training results of Wise and Olson (1995) and the 1- and 2-year follow-up data tell the following story. The PA group showed a clear advantage at the end of training for gains in phoneme awareness and phonological decoding. Moreover, the PA group's advantage in phonological decoding was significant 1 year after training in both the 1-year and 2-year follow-up groups. The PA group also showed a continued significant advantage in phoneme deletion 1 year after training only for the 2-year follow-up sample; the difference was not significant in the 1-year follow-up sample. According to Share's (1995) self-teaching hypothesis, the PA group's advantage in phoneme awareness and phonological decoding should have supported the more rapid growth of word recognition over that period.

Although the PA group showed significant advantages on the end-of-training individualized tests of words previously targeted in the ROSS stories, and a nearly significant advantage on the end-of-training PIAT test, the 1- and 2-year follow-up tests showed no significant treatment differences on the PIAT, WRAT, or time-limited word-recognition tests. Moreover, it is clear from the very small effect sizes for treatment differences at follow-up (–.16 to .25), that the reduced statistical power of the smaller follow-up samples is not the reason for the lack of significant treatment differences in word recognition. Statistical power was sufficient to detect the continued PA group advantages in phonological decoding and phoneme deletion 1 year after training.

Why did the PA group's continued significant superiority through the 1-year follow-up in phonological decoding and phoneme awareness fail to produce superior word recognition at the 1- and 2-year follow-up tests? We consider several possible answers. One we can emphatically reject is that the problem is in computer-based training. Olson, Wise, Johnson, and Ring (in press) reviewed several recent studies that compared different well-structured programs delivered both by classroom teachers and by individual tutoring without computers. Although the contents of the training conditions in these studies and the participant samples were not exactly like those used in this study, they did compare programs that varied in their relative amount of emphasis on phoneme awareness and phonological decoding (Brown & Felton, 1990; Foorman, Francis, Shaywitz, & Shaywitz, in press; Lovett, et al., 1994; Torgesen, Wagner, & Rashotte, this issue). The more explicit phonologically based programs produced better phonological decoding, but not significantly better word recognition in most measures at the end of training.

The research by Torgesen et al. (this issue) is most comparable to ours in its use of the Lindamood ADD training program (Lindamood & Lindamood, 1975) for its most phonologically based condition, and an EP program that focused more on reading text. One study used intense, one-on-one tutoring with reading-disabled children between 8 and 10.5 years of age for 80 to 90 hr (compared to our 25 hr) over an 8-week period. There were significant end-of-training differences favoring the ADD program in phonological decoding and phoneme awareness, but not word recognition. Follow-up data have not yet been reported for this study.

Another study reported by Torgesen et al. (this issue) used one-on-one tutoring to train children initially selected for low phoneme awareness, for an average of 88 hr from the second half of kindergarten to the end of second grade. The pattern of end-of-training effect sizes for the ADD over the EP group (computed from Table 2 in Torgesen et al.) was similar to the PA group over the CS group in our study: The effect sizes were .87 for a measure similar to our phoneme deletion, .97 and .91 for two measures of nonword reading, and .42 for the Woodcock test of word recognition. Our follow-up results suggest the value of testing these children 1 and 2 years after training to see if the significant advantages in phonological skills are maintained, and if this eventually leads to more rapid growth in word recognition.

Our results suggest that much more than 25 hr of training is required for most disabled readers to consolidate, automatize, and transfer their improved phonological skills to classroom and home reading so they may contribute to continued rapid growth in word recognition after intervention stops. The gains in standard scores of 7.5 to 10 points on the PIAT and WRAT tests during this limited training were encouraging because they indicate the children moved up relative to their peers' growth in word recognition. The improved standard scores remained stable on the WRAT but declined on the PIAT in the follow-up tests. We do not know why the WRAT and PIAT tests yielded such different results for the maintenance of standard-score gains, but even the WRAT results indicated that most children in the CS and PA groups were still well below the national norms (and thus even more below local norms) at the end of training and at follow-up. Many of these children still found themselves at a disadvantage relative to their classmates at the end of training, so they would have continued difficulty handling class reading requirements, and they probably read less both in and out of class.

There is much evidence that most children with reading difficulties tend to read less than good readers. This reduced print exposure may be both a cause and consequence of poor reading. Much of the increase in word-recognition standard scores during our training was likely due to the increase in structured reading practice required in both the CS and PA programs. When our training stopped, many participants' reading practice may have reverted to the lower levels that existed prior to their participation in the study, or they had less support for their difficulties in word decoding, or both. Classroom reading in Boulder tends to encourage unmonitored silent reading and making reasonable guesses from context, not the kind of structured and accurate decoding support the training and the computer had provided.

In addition to noting the need for increased training time, Wise and Olson (1995) suggested that the content of their PA program should be reinforced in several ways. One is to improve participants' flexibility in phonological decoding. Children who inflexibly apply their phonological decoding skills will often misread visually unfamiliar exception words that do not follow the most common grapheme–phoneme correspondence rules. A similar problem arises with inflexible dependence on regular grapheme–phoneme correspondences when reading visually unfamiliar multisyllabic words, where the correct vowel pronunciation and stress pattern may depend on lexical knowledge. Wise and Olson (1995) reasoned that if participants were more explicitly trained to try various vowel pronunciations when encountering unfamiliar exception and multisyllabic words, they would be more likely to recognize the correct decoding, assuming the word was in their oral vocabulary. A new version of our PA program will provide more explicit training in the flexing of vowels to find the correct decoding for exception and multisyllabic words. However, oral–lexical knowledge of visually unfamiliar exception and multisyllabic words is necessary for the flexing process to converge on the correct decoding. Some decoding failures in disabled readers may be due in part to a weak oral vocabulary, which may in turn be caused in part by less exposure to age-appropriate reading material. A third suggested improvement was to increase speed in phonological decoding. Although training effects on accuracy were strong in the PA condition, the disabled readers' phonological decoding may have been too slow to be readily employed in the fluent reading of text.

These results support a common observation in the field of learning disabilities: Children, especially those with learning disabilities, will learn what they are trained to do. The PA program included much training in phonological encoding and decoding with regular words and nonwords, but it included no specific training or practice in exception words (outside of reading the ROSS stories), in fluency, or in explicitly how to apply and practice the skills away from the computer. The study showed what children gain during training from explicit training in phonological skills, and the results there are strong and impressive. Interest may now be shifting to how to maintain gains after training, rather than to comparing differential gains during carefully controlled (if limited) training. To accomplish this, remediation studies will need to include specific training, practice, and reinforcement for learning to apply and extend the skills to everyday reading and writing away from the training setting.

Our current research goals include greater improvement of accuracy, flexibility, fluency, and application of phonological decoding away from the computer so that it can effectively serve as a self-teaching mechanism, leading to higher posttraining growth rates in word recognition. We are also exploring the specific contributions of the articulatory training, phoneme manipulation, spelling, and reading components of the PA program. Another research goal is to explore potential interactions between participants' individual processing profiles at pretest, and gains in the training conditions. In subsequent years, our PA and CS samples will become large enough to have the statistical power to detect interactions between individual variables and treatment condition.

Our final comments emphasize the advantages of computer-based research on training methods, and on instruction and remediation efforts. Many field comparisons of remedial programs are confounded by differences among the particular teachers or schools, or both, used in different training conditions, variables that can have big effects on training results. Comparing computer-based programs using the *same trainers in the same schools can avoid these potential confounds and can help* ensure that specified training regimes are followed with equal enthusiasm and time on task.

Results with both the PA and CS programs clearly show the benefits and economy of computer-based remediation for reading disabilities. The benefits are clear in the rapid growth rates for word recognition during the limited 25-hr training presented over 4 months in the research studies. The fact that the rate of growth diminishes after training suggests that more training and support is necessary to help these growth rates become stable. These benefits should continue to accrue with longer training over several semesters to bring most children with reading disabilities into the local normal range. The need for long-term individual support highlights the significant economy of computer-based remediation. Intense, individual support with tutors in programs such as Reading Recovery is extremely expensive (Shanahan & Barr, 1995). The use of well-designed computer programs and a trainer in small groups of disabled readers can provide more individualized support at much less cost. Continued research should help guide the development of programs that can more efficiently identify and remedy the specific processing deficits of individual children with reading problems.

ACKNOWLEDGMENT

The research was supported by NICHD grants HD 11683 and HD 22223.

We thank the principals, staff, and students from four Boulder Valley schools who participated in this research; the trainers John Green, Sally Moody, Beverly Peterson, and Kate Wise; programmer Jennifer Restrepo; and research assistants Laura Kriho and Bonnie Houkal. We thank the consultants Pat Lindamood and Beverly Peterson for help with the ADD training, and Leigh Kirkland and Michael Meloth for help with Reciprocal Teaching. Barbara Foorman and two anonymous reviewers provided helpful comments on an earlier version of the article.

REFERENCES

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

Dunn, L. M., & Markwardt, F. C. (1970). *Examiner's manual: Peabody Individual Achievement Test.* Circle Pines, MN: American Guidance Service.

Foorman, B. R., Francis, D. J., Shaywitz, S. E., & Shaywitz, B. A. (in press). Early interventions for children with reading problems: Study designs and preliminary findings. In B. Blachman (Ed.), *Foundations of reading acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Jastak, J., & Jastak, S. (1978). *The Wide Range Achievement Test–Revised.* Wilmington, DE: Jastak Associates.

Jorm, A. F., & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics, 4*(2), 103–147.

Lindamood, C., & Lindamood, P. (1975). *Auditory discrimination in depth.* Columbus, OH: Science Research Associates Division, Macmillan/McGraw-Hill.

Lovett, M. W., Borden, S. L., DeLuca, T., Lacerenza, L., Benson, N. J., & Brackstone, D. (1994). Treating the core deficits of developmental dyslexia: Evidence of transfer of learning after phonologically- and strategy-based reading training programs. *Developmental Psychology, 30*(6), 805–822.

Olson, R. K., & Wise, B. W. (1992). Reading on the computer with orthographic and speech feedback: An overview of the Colorado remediation project. *Reading & Writing, 4*(2), 107–144.

Olson, R. K., Wise, B. W., Conners, F., Rack, J., & Fulker, D. (1989). Specific deficits in component reading and language skills: Genetic and environmental influences. *Journal of Learning Disabilities, 22,* 339–348.

Olson, R. K., Wise, B., Johnson, M., & Ring, J. (in press). The etiology and remediation of phonologically based word recognition and spelling disabilities: Are phonological deficits the "hole" story? In B. Blachman (Ed.), *Foundations of reading acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition & Instruction, 1*(2), 117–175.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27,* 28–53.

Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly, 30,* 958–996.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*(2), 151–218.

Vellutino, F. R., Scanlon, D. M., & Denkla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 639.

Wechsler, D. (1974). *Examiner's manual: Wechsler Intelligence Scale for Children–Revised.* New York: Psychological Corporation.

Wise, B. W., & Olson, R. K. (1992). How poor readers and spellers use interactive speech in a computerized spelling program. *Reading & Writing, 4*(2), 145–163.

Wise, B. W., & Olson, R. K. (1995). Computer-based phonological awareness and reading instruction. *Annals of Dyslexia, 45,* 99–122.

Wise, B. W., Olson, R. K., Anstett, M., Andrews, L., Terjak, M., Schneider, V., Kostuch, J., & Kriho, L. (1989). Implementing a long-term computerized remedial reading program with synthetic speech feedback: Hardware, software, and real world issues. *Behavior Research Methods, Instruments, and Computers, 21,* 173–180.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23,* 159–177.

Manuscript received October 25, 1996
Final Revision received February 3, 1997
Accepted February 10, 1997