Reading on the computer with orthographic and speech feedback

An overview of the Colorado remediation project

RICHARD K. OLSON and BARBARA W. WISE
University of Colorado at Boulder

ABSTRACT. In this paper we present an overview of a computer program directed toward the remediation of children's deficits in word recognition and phonological decoding. In the present studies, 138 children read stories on the computer, in their school, for a half hour per day during a semester. Children were trained to request synthetic-speech feedback (DECtalk) for difficult words by targeting the words with a mouse. Different groups received whole-word feedback, wherein targeted words were highlighted and spoken as a unit, or segmented feedback, wherein segments of words (onsets, rimes, or syllables) were sequentially highlighted and spoken by the computer, requiring the child to pay attention to and blend the segments. Both whole-word and segmented feedback resulted in almost twice the gains in standardized word recognition scores compared to control groups that spent an equal time in their normal remedial reading program. Most important, the computer-trained groups improved their phonological decoding of nonwords at about four times the rate of the control group. However, there was a significant interaction between level of deficit severity and optimal feedback condition. The most severely disabled readers showed the largest phonological decoding gains from syllable feedback, while the largest gains for the less severely disabled readers were from onset-rime feedback. The disabled readers' level of phonological awareness at pre-test was the strongest predictor for gains in word recognition and phonological decoding. Implications of the results for future training programs are discussed.

KEYWORDS: Computer-based remediation, Individual differences, Phonological awareness, Reading disability, Synthetic-speech feedback

INTRODUCTION

The recent development of high-quality synthetic speech for microcomputers has enabled promising new approaches to the remediation of reading disabilities. In this paper we present an overview of a computer-based research and training program that has been conducted at the University of Colorado beginning in 1985. Initial studies validated the use of synthetic-speech feedback for learning difficult words in stories and word lists presented on the computer (Olson, Foltz & Wise 1986; Wise 1987). Subsequent research has focused on the long-term training of generalizable phonological decoding skills. Second-to sixth-grade reading-disabled children in the schools have read stories on the computer for a half hour each day. Children were trained to target difficult words to obtain immediate orthographic and speech feed-

Reading and Writing: An Interdisciplinary Journal **4**: 107—144, 1992.

back. A primary research goal has been to compare the benefits of different levels of segmentation in orthographic and speech feedback for targeted words. For some children, targeted words were highlighted and spoken as a whole-word unit (e.g., cupcake). Other children received feedback segmented into syllables (e.g., cup/cake) or into onset and rime segments (e.g., c/up/c/ake) for each targeted word. The sub-word units were sequentially highlighted and spoken by the computer. Pre-test to post-test gains over a semester in word recognition and phonological decoding have been compared across the different feedback conditions. Gains of the computer-trained children have also been compared with those of matched control groups that instead received their normal course of instruction in remedial reading or language-arts classes.

We begin this overview with a brief discussion of behavioral-genetic studies at the University of Colorado that have provided evidence on the nature and etiology of specific reading disabilities and have helped guide the design of the training studies. In the second section, preliminary studies of synthetic speech intelligibility and different computer hardware configurations are reviewed. The third section focuses on the theoretical rationale and research that guided our selection of different methods for segmenting words into syllables and sub-syllable units. The fourth and major section of the paper reviews the design and results of two long-term training studies in the schools.

I. IMPLICATIONS FOR REMEDIATION FROM THE COLORADO READING PROJECT

Our research on the computer-based remediation of reading disabilities has been motivated and informed by an on-going study of their genetic and environmental etiology. In the Colorado Reading Project (DeFries et al. 1991), reading-disabled identical and fraternal twins from the third to the twelfth grade are ascertained from school records and studied extensively in the laboratory. At least one twin in each pair must have a school record indicating reading problems, and when tested in the laboratory, must be below about the tenth percentile (based on local norms) for a composite of reading, spelling, and comprehension scores from the Peabody Individual Achievement Test (PIAT, Dunn & Markwardt 1970). Additional criteria include a verbal or performance IQ score above 90, English as a first language, no obvious sensory or neurological problems such as seizures, and no obvious educational constraints such as poor school attendance. These criteria are similar to those used in our studies of computer-based remediation in the Boulder schools.

Disabled readers selected under the above criteria display performance profiles for reading related measures that have direct implications for remediation. The disabled readers tend to have greater difficulty in reading isolated words and spelling than in the comprehension of oral or even written text (Conners & Olson 1990; Stanovich 1986). This profile suggests that much of the disabled readers' deficit in reading comprehension is caused by their difficulties in recognizing or decoding printed words (Perfetti 1985). Of course, some disabled readers, particularly those with low general intelligence and poor educational background, may also suffer from other cognitive deficits that constrain their comprehension of oral and written text. A number of successful training programs have been directed toward the remediation of higher-level comprehension processes (see Dole et al. 1991, for a review). However, these programs will have limited success when basic word-recognition processes are deficient. Therefore, while our computer-based training project monitors and encourages reading comprehension, it focuses primarily on the development of word recognition and related component processes.

Phonological decoding, commonly measured by the oral reading of nonwords (e.g., 'tegwop'), is a particularly important component processes in word recognition. Results from the Colorado Reading Project have provided some of the strongest evidence that phonological decoding and related phonological awareness skills are substantially lower in most disabled readers than expected from their level of word recognition (Olson 1985; Olson et al. 1989; Rack, Snowling & Olson, in press). Moreover, behavioral-genetic analyses of twin data have revealed that phonological decoding deficits are highly heritable, account for most of the heritable variance in word recognition, and may originate from heritable deficits in phonological awareness (Olson et al. 1989; Olson et al. 1991).

Deficits in phonological decoding constrain the development of disabled readers' word recognition in two ways. First, attempts to sound out an unfamiliar word may often be incorrect. If the error is not recognized and corrective feedback is not given, as is often the case for silent reading in the classroom, the wrong print-to-sound association is reinforced. This incorrect association may subsequently interfere with memory for the relation between print and sound in other similarly spelled words (Jorm & Share 1983). Second, disabled readers typically require more learning trials with corrective feedback than normal children to establish an automatized level of recognition for a printed word (Reitsma 1983). When disabled readers look at an unfamiliar word and are told what it says, their encoding of the relation between the word's orthography and phonology may be limited by deficits in associating units in orthography and phonology smaller than the word, the essence of phonological decoding.

For the above reasons, the remediation of disabled readers' phonological decoding deficits is a primary goal of our computer-based training program. Some disabled readers ultimately develop strong word-recognition skills through extensive reading practice and print exposure, but their phonological-decoding skills remain deficient and they have great difficulty decoding new words without assistance (cf. Campbell & Butterworth 1985). Improved phonological-decoding skills would enable children to decode difficult words successfully without external feedback, support the more efficient development of automaticity in word recognition, and free up additional resources for reading comprehension. Although many disabled readers' phonological-decoding deficits are strongly influenced by genetic factors (Olson et al. 1989, 1991), this only implies that some extraordinary environmental intervention may be required, beyond what most reading-disabled children currently receive in their homes and schools.

Most children learn to read effectively from the various reading instructional methodologies used in the schools. They become able to read stories of interest to them, and notice enough print and sound regularities and analogous words to be able to figure out new words and understand text with the amount of support available to them in a typical 30-children classroom. However, some children progress at a frustratingly slow pace or progress not at all with the same amount of support. These reading-disabled children might benefit from the opportunity to read interesting stories with intense and precisely controlled one-on-one feedback for word-decoding problems. Unfortunately, limited resources in the schools and parents' time and/or parents' limited reading skills often constrain the availability of effective tutoring for the additional needs of disabled readers. Even when help for word-decoding problems is available, many disabled readers may not ask for it because of inconvenience or embarrassment. Therefore, we wondered if an effective talking-computer program could be designed to motivate reading and provide feedback for the development of disabled readers' word-recognition and phonological-decoding skills.

II. INITIAL STUDIES OF TEXT-TO-SPEECH SYSTEMS AND HARDWARE CONFIGURATIONS

DECtalk, a stand-alone text-to-speech synthesizer for microcomputers, was first released by the Digital Equipment Corporation in 1984. An earlier implementation of text-to-speech synthesis (Echo, Street Electronics) had been available for the Apple II, but we judged that its intelligibility was not adequate to provide useful feedback for disabled readers (later confirmed by Wise et al. 1989). DECtalk's high intelligibility, even for words isolated from context, was verified by a comparison with normal human speech (Olson et al. 1986). A group of reading-disabled children correctly repeated 95% and a normal adult group repeated 96% of 120 unrelated words spoken by DECtalk. The respective group scores were 98% and 96% for recorded human speech. The comparison indicated that DECtalk's intelligibility was clearly adequate to provide feedback for targeted words in context and for other instruction in the training program. The main limitation of DECtalk for the educational market has been its cost, currently US$ 1600 to non-profit institutions. The price is expected to decrease to about US$ 840 with educational discounts when the system is implemented on a single IBM PC-compatible board that will be available in early 1992.

It is beyond the scope of this paper to present an extensive review of different options for speech feedback, but two general points should be mentioned. First, the basic technology developed at MIT by Dennis Klatt and incorporated in DECtalk is available in two other systems. A Swedish company (Infovox) and Speech Plus (Berkeley, California) have both produced high-quality but similarly expensive text-to-speech synthesizers on IBM PC-compatible boards. Competition is likely to lower the price for these systems as the educational market expands. In the future, the power and multitasking capabilities of the newer personal computers may support software-based programs for high-quality synthetic speech, further lowering the price of this technology.

The second general point about computer-speech technology is the changing relative advantages of synthetic speech versus digitized human speech. At the time of our last review, the high memory demands of digitized speech limited its usefulness (Olson & Wise 1987). Now, the rapidly declining cost of magnetic and CD ROM memory has made digitized human speech a more viable option for speech feedback in reading. A Canadian company (Discis Knowledge Research, Toronto) is currently marketing story books on CD ROM disks for MAC computers. In addition to excellent graphic displays of the story pictures, the disks store digitized speech for each word that can be requested by the reader. Newer Apple computers support the creation and reproduction of digitized speech, and digitizing boards with data compression algorithms are available for the IBM PC from Street Electronics. Both synthetic and digitized speech are likely to play important roles in computer-based reading instruction, but high-quality synthetic speech still has a clear advantage in its ability to pronounce novel character strings, as in the spelling program described by Wise and Olson (this issue).

DECtalk was first used to provide feedback for word decoding problems in a study conducted by Olson et al. (1986). Disabled readers read stories on the computer over two one-hour sessions. In one session, they targeted difficult words in the stories, which were immediately highlighted in reverse video and spoken by the computer. In the other session, difficult words were targeted and highlighted but not spoken by the computer. Speech feedback provided clear advantages for learning targeted words and improving subjects' comprehension of the stories. The children's general reaction to the system was very positive and most said they would like to continue reading on the computer with speech feedback. The results of this pilot study encouraged us to apply for funding from the National Institutes of Health for the development of a long-term training program.

Several hardware decisions were made as a result of our initial study (Olson et al. 1986). Some subjects used a mouse while others used a light pen to target words. Performance was not significantly different with the two devices. The mouse required only a short period of practice for effective use.

The light pen required no initial practice, but the light pens were less reliable, and they scratched the screen. The mouse has been used in all subsequent studies. A second hardware decision was whether to use the Apple or IBM PC computers. Both systems were used in the pilot study. At that time, IBM PC/XT clones were considerably less expensive and easier to program than the Apple systems. Our subsequent studies have used PC/XT clone systems with serial ports for the DECtalk and mouse, a 40 MB hard disk, 360K floppy, 640K main memory, a Hercules graphics video adaptor, and monochrome monitor. The programs have also been adapted for the newer 386SX systems with VGA color monitors.

III. THEORETICAL AND EMPIRICAL BACKGROUND FOR METHODS OF SEGMENTATION

Different subjects in the Olson et al. (1986) study received orthographic and speech feedback that was presented in whole-word, syllable, or sub-syllable units. The subject samples were small in each group and there were no significant differences between levels of segmentation for the short-term learning of targeted words. However, we hypothesized that long-term training effects on word recognition and phonological decoding would vary across the different levels of feedback segmentation.

Whole-word feedback should certainly help children learn the specific words they target. From some theoretical perspectives, it might be the best type of feedback for the development of reading skills. Some 'Whole-Language' theorists (e.g., Goodman 1967) suggest that children ought to guess the identity of unfamiliar printed words through story context and the initial sound of the word. Any more detailed phonological decoding process and attention to word segments is thought to distract the child from the most important higher-level linguistic processes in reading. Goodman would not agree that even whole-word feedback should be provided because he has argued against any correction of decoding errors that make sense within the context. Some clinicians have suggested avoiding the training of phonological decoding processes for the majority of disabled readers whose phonological decoding is very weak (e.g., Boder & Jarrico 1982). They argue instead that such children should be encouraged to use their greater strength in reading words as whole units.

In addition to learning targeted words, whole-word feedback might also be expected to improve disabled readers' generalizable phonological decoding ability. Connectionist theorists have demonstrated that computer programs can learn relations between sub-word orthographic and phonological units even though the orthographic and phonological-feedback units presented to the system are all whole words (Seidenberg & McClelland 1989). Besner, Twilley, McCann and Seergobin (1990) noted that the phonological decoding of this simulation was really not very good, but we are impressed that some significant phonological decoding skills did develop in the whole-word simulation. Seidenberg and McClelland suggest that their simulation may reflect the normal process of development for children's phonological decoding. However, disabled readers usually fail to develop adequate phonological decoding and phonological awareness with standard reading experience. Therefore, we hypothesized that they might show greater gains from feedback explicitly segmented into units smaller than words. Several levels of segmentation could be used. The syllable is a natural unit that is accessable even to children who have very poor phonological awareness (Fox & Routh 1980; Rozin & Gleitman 1977). Syllables can also be divided into individual phonemes or some intermediate sub-syllabic units. Depending on their phonological awareness and severity of reading deficits, disabled readers might vary in their optimal level of feedback segmentation.

We now turn to theory and data that guided our initial selection of rules for segmenting words. Before designing our segmentation conditions for the long-term training studies, it was necessary to determine optimal ways of defining syllable and sub-syllable segments within words. In three short-term studies, children in the first and second grades learned lists of words on the computer with the help of segmented orthographic and speech feedback. The first study addressed two different ways of defining sub-syllable segments. The second addressed two different ways of defining syllables. The third compared whole-word, syllable, sub-syllable, and phoneme segmentation.

A. *Sub-syllable segmentation*

The first set of experiments compared two types of intermediate sub-syllable units larger than the phoneme (Wise 1987; Wise, Olson & Treiman 1990). An 'onset-rime' segmentation condition divided syllables between the initial consonant cluster and the vowel-consonant group (e.g., d/ish, cl/ap). The other condition divided the syllable immediately following the vowel (e.g., di/sh cla/p). Treiman (1983) had shown that division at an onset-rime boundary was easier for adults and children to use in spoken language games, and we hypothesized that this segmentation would also be more helpful in feedback for learning to read words.

The two types of sub-syllable segments were compared in four orthographic types of words (CCVC and CVCC with 3 or 4 phonemes) to see whether one type would benefit word learning more than the other. Twenty first-graders were pretested for initial knowledge of a list of words. They were then given three training trials per word; the computer highlighted the orthographic segments and gave concurrent speech feedback for the segments when the child touched the word on the screen with a light pen. The child then attempted to blend the segments and say the word aloud, without feedback from the experimenter. After training on all words in the list, subjects were post-tested without segmentation or speech feedback. The results showed significant advantages in post-test recognition for words that were segmented at the onset-rime boundary (e.g., d/ish).

B. *Syllable segmentation*

Wise (1987) studied two different ways of segmenting words into syllables using the same computer-speech methodology as in the above sub-syllable study. Twenty normal second-grade readers were trained in their schools. Syllabic division based on the 'vocalic center group' promoted by Spoehr and Smith (1973), (e.g., rea/der, gar/den), was compared with syllabification by Taft's (1979) 'BOSS' rules, based on morphemic and 'orthotactic' considerations, which always preserve morphemic units in multimorphemic words (e.g., read/er). Taft argued that readers form units in reading by proceeding left-to-right through the vowel group, and continuing on until reaching either a morpheme boundary (e.g., read/er), a new pronounced vowel group (e.g., gard/en), or an orthotactically illegal letter combination (e.g., nap/kin). The 'BOSS' segmentation condition for defining syllables appeared to have led to a significant advantage on post-testing compared to the vocalic center group condition.

C. *Levels of segmentation*

Based on the results of the above two studies, 4 levels of segmentation were compared for helping children learn single and multisyllabic words: (1) whole-word (e.g., reader), (2) 'BOSS' type syllables (e.g., read/er), (3) onset-rime sub-syllables (e.g., r/ead/er), and (4) grapheme-phoneme units (e.g., r/ea/d/er) (Wise 1987; Wise, in press). Subjects included 56 first-graders and 56 second-graders divided into groups of low and average reading ability. The training procedure in this experiment was quite similar to the above studies, except that it occurred over a 2-day training session.

In this short-term study, whole-word and syllable feedback proved to be the most helpful for first and second graders of low and average reading ability. These two conditions did not differ significantly from each other. Onset-rime segmentation proved somewhat less helpful on multisyllabic words for the younger and low reading groups, but their gains were still substantial and significant. For monosyllabic words, onset-rime segments proved as easy to blend and as effective in benefiting word recognition as the whole-word units, even for the younger and lower ability readers.

The strongest, most consistent, and most significant result was that grapheme-phoneme segmentation aided word learning the least for all groups. The main problem was with difficulties in blending during training. For example, even after three training trials, low ability second graders could correctly blend the phoneme segments for only 49% of the words they had misread on a pretest, compared to 78% for onset-rime and 96% for syllable segmentation conditions.

Poor learning from grapheme-phoneme feedback was also found by Spaai, Reitsma & Ellerman (1991) in a study of Dutch children who had about 9 months of reading instruction. Spaai et al. used digitized speech to compare whole-word and grapheme-phoneme feedback for learning a list of words. Learning in the grapheme-phoneme condition was significantly lower than from whole-word feedback and not significantly different from a control group that received no feedback.

The problem with isolated phoneme presentation may be related to the observation by A. M. Liberman et al. (1967) that phonemes sound different in contexts, and information about the segments overlaps to a considerable degree. Although single phoneme segmentation and blending is a common educational practice for beginning readers, I. Y. Liberman (1983) has pointed out that it may be counterproductive to try to isolate many phonemes in this way, especially the stop consonants p, b, t, d, g, and k, which cannot be pronounced without adding a neutral vowel sound. Also, any word divided into phonemes will necessarily have more units than the word divided into onset/rime units or syllables, so memory limitations may also be causing difficulty in this condition.

We concluded from the above results that it would not be useful to use feedback segmented entirely by individual phonemes in our long-term studies. However, the other three conditions led to greater benefits in word learning. Although whole-word and syllable feedback benefitted blending and learning slightly more than onset-rime feedback in the short-term study, the presence of context in stories might help in the successful blending of onset-rime segments. Therefore, we hypothesized that onset-rime feedback might be best for the long-term development of disabled readers' generalizable phonological coding skills.

IV. LONG-TERM TRAINING STUDIES

The main focus of our long-term training studies has been on disabled readers' development of generalizable phonological decoding skills as well as their learning of specific words targeted while reading stories on the computer. Two studies (Phase I and Phase II) are discussed in this section. Phase I was conducted in the spring semester of 1988. This study involved a high level of experimenter monitoring and encouragement for children reading on the computer. In spite of the fact that subjects averaged only 6.4 hours actual reading time on the computer over the semester, children with segmented feedback demonstrated impressive pre-test to post-test gains in phonological decoding, in contrast to small gains in the whole-word feedback and untrained control conditions. This study, previously reported in Wise et al. (1989), is reviewed again in this report because there were notable changes in the results of a subsequent study (Phase II). The second study included essentially the same methodology as Phase I, except that the sample was much larger and subjects received substantially less experimenter pretraining and monitoring for their targeting of difficult words. In spite of the longer reading time in Phase II (8.1 hrs.), the segmented feedback conditions showed reduced but still significant gains in nonword reading compared to the control group. Most important, gains from whole-word feedback, which had not been significantly better than the control group in Phase I, were significantly better in Phase II, and onset-rime feedback was the strongest trained condition in Phase I, but nonsignificantly the weakest trained condition in Phase II. The different results for onset-rime feedback across the studies will be partly explained by differences in sample characteristics.

In the first section (A), we present an overview of the subject characteristics and general methodology common to both studies. In the second section (B), differences between Phase I and II in training, sample characteristics, and results are presented. The third section (C) includes an analysis of the effects of different levels of reading-deficit severity on gains in phonological decoding within the combined sample from Phase I and II. Sections (D) and (E) focus on the differences in disabled readers' gains in phonological decoding and word recognition as a function of their pre-test scores for phonological awareness.

A. *General methodology*

1. *Program development*

The long-term training studies required the development of programs to administer pre- and post-tests, present texts and feedback, and monitor subjects' performance. A large number of short stories and books were entered on the computer to provide enough reading material for primary to sixth-grade ability levels. Multiple choice comprehension questions were written to be inserted after every 5 to 10 pages of text. As new reading materials were entered, each new word was coded into a dictionary that stored segmentation and pronunciation information for all feedback conditions. The dictionary now holds over 20,000 words. A more detailed description of the programs is given in Wise et al. (1989).

2. *Subject selection and assignment to conditions*

Prospective subjects were referred by teachers based on poor classroom performance in reading and on various standardized tests of reading given in the schools. They were then given the PIAT word-recognition test by our experimenters and the ratio of grade-level performance on the PIAT to the subject's performance expected by school-grade was computed. The average PIAT-grade/school-grade ratio score in the Boulder schools is about 1.2, reflecting the relatively high socioeconomic and educational level in this area. Therefore, although a number of the trained and control subjects had PIAT scores approaching the national norm, they were relatively poor readers in the Boulder schools. We decided to include a broad range of reading-deficit severity in the sample to observe the interaction of severity with benefits from the different training conditions. Additional selection criteria for the disabled and normal subjects included a normal-range IQ (at least 90 on verbal or performance subscales of the WISC), no obvious sensory or neurological deficits such as seizures, normal school attendance, and English as a first language. Phase I and II sample characteristics for age, school grade, percentage of males, the ratio of grade-level performance on the PIAT to age-expected performance, and IQ are presented in Table 1.

Table 1. Sample characteristics for subjects in Phase I and II*

Phase	Age years	Grade	% Males	PIAT ratio	WISC-R IQ
Phase I	10.1	3.9	60%	0.81	104.0
(N = 37)	(1.2)	(1.1)		(0.25)	(4.2)
Phase II	9.8	3.8	62%	0.78	103.3
(N = 149)	(1.0)	(0.9)		(0.23)	(11.4)

* Standard deviations in parentheses.

Disabled readers in grades 2—6 were given pre-tests at the beginning of the semester that included a second measure of word recognition and a measure of phonological decoding (see below). Then the subjects were randomly assigned to the whole-word, syllable, or onset-rime training conditions or the untrained control condition, except that mean levels of school grade and phonological decoding were made as equivalent as possible across the conditions.

3. *Pre- and post-tests*

(a) *Peabody individual achievement tests* (PIAT) *for word recognition, reading comprehension, and spelling* (Dunn & Markwardt 1970). The word-recognition sub-test of the PIAT contains 66 words of increasing difficulty in rows across a page. This test was adapted for presentation on the computer by presenting each successive word in isolation on the screen and having the experimenter score the final response as correct or incorrect. As in the standardized PIAT, the computer test was stopped when the subject missed 5 of the last 7 items. Test-retest reliability for the booklet version of the PIAT word recognition test was reported to be 0.89. In addition, the PIAT standardized comprehension and spelling tests were administered in their booklet format. The comprehension test requires subjects to point to one of four pictures that best represents the meaning of a printed sentence. The spelling test requires subjects to choose the correct spelling for a spoken word from four printed alternatives. The PIAT tests are standardized by grade level according to national norms. the reported reliabilities of the comprehension (0.64) and spelling (0.65) tests are rather low for assessing gain scores in the limited ability range and training time of the present study. The measures were included because their short test time had minimal impact on the time left for computer training.

(b) *Timed word-recognition.* A timed word-recognition test developed in our laboratory was included for increased sensitivity to changes between pre- and post-tests and for a measure of subjects' ability to read words under a limited exposure duration. The timed word-recognition test included a larger number of items (220) than the PIAT (66), particularly in the ability range of our subjects. The words were presented in order from the least to most difficult items. The subjects were initially given a 14 item screener test to place them at an appropriate difficulty level within the list and avoid the reading of many words that would be too easy. Each word was presented for two seconds on the computer. The subject's final response to each word was scored by the experimenter and the next trial was presented. The test was stopped when the subject made errors on five out of the last seven trials. The score used in the present analyses was the number of words read correctly up to the stopping point in the test.

(c) *Phonological decoding* (*oral nonword reading*). Students attempted to pronounce eighty-six pronounceable nonwords of widely varying difficulty, ranging from one-syllable items with two phonemes and three letters to two-syllable items with eight phonemes and nine letters. The nonwords were presented individually on the computer, in random order of difficulty, over two days in 43 trial blocks. The subject's last response was scored as correct or incorrect by the experimenter and the next trial was initiated. The subject's score was the percent of nonwords read correctly.

(d) *Phonological awareness in language: Pig latin.* Subjects in Phase I and the fall semester of Phase II were instructed by the computer on how to play 'Pig Latin' with real words spoken by DECtalk. Instructions demonstrated how to take off the sounds before the rhyming part of the word and begin the new word with the rhyming part (e.g., pig = ig). Then the beginning sounds plus 'ay' were added to the end of the new word (e.g., pig = igpay). The computer gave four demonstration and five practice items. The test itself consisted of 44 items in order of least to most difficult items. The DECtalk pronounced the target word, used it in a sentence, then pronounced the word again. The experimenter scored the student's last response and the next trial was presented. All students did at least the first 28 items, and after that stopped whenever they missed 6 out of the last 7 items. The subject's final score was the percent of correct responses of the total 44 items, regardless of whether they actually completed all trials.

(e) *Phonological awareness in language: Phoneme deletion.* Subjects in Phase II listened to tape-recorded nonwords through headphones. The recording pronounced a nonword and then pronounced a phoneme to be deleted from the nonword. (e.g., 'Say barp, without the /puh/'). Instructions on the tape included 4 demonstration items. The student had 6 seconds to respond before the next item was presented. The tester scored the student's last response as correct or incorrect. In Fall 1988 semester of Phase II, there were 58 test items of increasing difficulty. Testing was discontinued if the student made no correct responses among the first 15 items, or made less than three correct responses among the first 26 items. The test was changed for the Spring 1989 semester of Phase II. The new test presented the nonwords differently. The recording now asked the student first to repeat the nonsense word, and then to say it without one of the phonemes ("Say barp", 2 sec interval). "Now say barp without the /puh/". This change was made to increase the likelihood that the student would hear the original nonword correctly. The test was also shortened after doing item analyses on earlier versions of the test. The test was reduced to 40 items divided into easy, core, and hard sections. Testing was discontinued if the student made no correct responses in the 8 item easy section. Otherwise the student completed the next 24-item core section. If the student responded correctly to one or more of 6 designated "difficult" items in the core section, they attempted the final 8-item difficult section. The subject's final score for both versions of the test was the percent of items pronounced correctly, based on the total number of items in the test.

(f) *WISC-R Full-Scale IQ* (Wechsler 1974). Full-scale IQ was estimated from administration of the Vocabulary, Similarities, Picture Arrangement, and Block Design subtests of the Wechsler Intelligence Scale for Children-Revised.

4. *Training procedure*

The general training procedure in our long-term studies was to have children read interesting stories on the computer for a half hour each day. Children were trained to 'target' words with a mouse to ask the computer for help with difficult words. Orthographic and speech feedback was given according to their assigned segmentation condition (see section (e) below).

(a) *Story selection.* Children selected stories from a menu of available titles in his or her assigned reading-level directory, which was initially determined by the child's pre-test word-recognition score and adjusted during training if necessary. We wanted the words in the stories to be easy enough for each subject so that reading would not be too disrupted by decoding problems, but difficult enough so that subjects would need feedback for 1—5 words per page. Each page might contain about 20 words in the primer directory, ranging up to 100 words in the higher directories. After the child selected a title from the menu, a short paragraph describing the selected story was presented on the computer. The child could ask the computer to read this description aloud. The child could then read that story or return to the menu for selection of another title.

(b) *Target pre-training and weekly monitoring.* During several initial training sessions, subjects read the stories aloud. The tester prompted them to target any miscue they had failed to correct by the end of the sentence, excluding short function words such as 'a' or 'the' that did not change the meaning of the sentence. This training was critical for many subjects who seemed to have little sensitivity to most of their decoding errors. Many of these subjects had previously been encouraged by their teachers to skip over or guess difficult words by using context clues. The subject's oral reading was monitored weekly by the experimenter to see if unknown words were consistently targeted, to shift subjects to higher or lower reading-level directories when necessary, and to improve sensitivity to decoding errors. The amount of initial training and weekly monitoring differed for Phase I and Phase II, as described in the next section. Our general goals for targeting behavior were first to have subjects consistently and spontaneously target a high percentage of words they missed while reading aloud for the experimenter, and second, to have subjects continue good targeting practices when reading independently on the computer. Our success in the second goal was inferred from the ratio of targeting frequency on independent days to targeting frequency on monitored days.

(c) *Comprehension monitoring and targeted word review.* Comprehension for recently read material and recognition of recently targeted words was monitored by the computer. Every 5—10 pages, at a natural pause in the story, the computer presented a multiple choice question concerning the immediately preceding material. Subjects could choose to read the question and alternative answers themselves, or could choose to have the computer read them. Then the subject selected one alternative answer with the mouse, and the computer scored the response and read the correct answer aloud. Following the comprehension question, the program presented five of the previously targeted words on the screen for recognition. If the subject had targeted less than five words, the program added some of the most difficult words from the text to make a 5-word test. During the independent training sessions, subjects tried to read each test word silently as its orthography was highlighted without speech according to the assigned segmentation condition. The computer then spoke the whole word, and subjects scored whether their unaided attempt was correct or incorrect. In monitored sessions, the subject read the words aloud and the experimenter scored the responses. After completion of the short comprehension and word recognition tests, subjects continued reading the story until the next test-break or end of the story.

(d) *Training conditions and untrained control condition.* The conditions were (1) whole-word feedback, (2) syllable feedback (by a combination of BOSS and stress rules), (3) onset-rime feedback (prefixes and two-letter syllables were not segmented), and (4) untrained control. When feedback for a word was requested in the first three conditions, the word or its segments were initially highlighted without speech and the subject was encouraged to attempt to decode the word. After a brief delay, the word or its orthographic segments were sequentially highlighted by reverse-video in synchrony with the speech segments. The segments were highlighted for an interval based on [(time × number of letters in segment) + 300 msec, where time = 300 msec. for whole word and 400 msec. for segmented conditions].

Control subjects were given the pre- and post-tests at the same time as the trained subjects, but they were not given any training on the computer. The control subjects received the normal course of reading instruction provided by the schools in their special education or language arts classrooms. In nearly all cases, the trained subjects' time on the computer replaced an equal amount of time in their normal course of language arts instruction which included a mix of reading, writing, and spelling.

B. *Differences in training, samples and results between Phase I and Phase II*

The first group of disabled readers in grades 3—6 was trained during the 1988 Spring semester (Phase I). A complete report of the Phase I study is presented in Wise et al. (1989). Subjects in Phase II were trained in the fall semester of 1988 and the spring semester of 1989. The basic methodology was the same in Phases I and II, but the results differed due to changes in subject pre-training, experimenter monitoring of targeting performance, and differences in deficit severity within the training conditions across phases. In this section we discuss differences in training procedures between Phases I and II and review their different results in targeting behavior, reading comprehension, and gains in nonword reading and word recognition. Section C focuses on the combined results of Phases I and II for analyses of subgroup by treatment condition interactions.

1. *Phase I and II differences in training and daily behavior*

In Phase I, a total of 27 reading disabled children read on the computer for at least 4 hours in either the whole-word, syllable, or onset-rime training conditions. Ten subjects were in the untrained control condition. A fourth feedback condition was included in Phase I that involved a combination of syllable and onset-rime feedback (see Wise et al. 1989). This condition is not discussed in this paper because there was no comparable condition in Phase II and it did not present any unique training benefits.

The mean reading time for the 27 trained subjects was 6.4 hours with no significant difference across feedback conditions. The mean total time on the system was 10 hours, which included selecting stories, listening to instructions, answering comprehension questions, and taking the intermittent comprehension and word recognition tests after every 5 to 10 pages. About half of the children's time on the computer was monitored by an experimenter, including the initial several sessions of pre-training for targeting difficult words, follow up monitoring of targeting behavior, and general encouragement.

After pre-training, the ratio of average targeting frequency on nonmonitored sessions to monitored sessions ('target ratio') was 27/33, or 0.82 (see

Table 2). The target ratio indicates how adequately subjects were targeting unknown words in independent reading sessions. The large standard deviation for target ratio (sd = 0.51) reflects the fact that a few subjects targeted about twice as many words when they were reading independently, reflecting their uncertainty about reading words correctly when there was no experimenter to provide feedback. There were no significant differences in target ratio across feedback conditions ($F(2,25) = 0.189$, $p > 0.05$). The intermittent comprehension questions averaged 95% correct in monitored and 90% correct in non-monitored sessions. Monitored word recognition tests for targeted words averaged 81% correct.

Table 2. Training time and daily performance on intermittent comprehension and word-recognition tests in Phase I and Phase II*

Phase	Reading time	Target ratio	Word M	Comprehension NM	Comprehension M
Phase I	6.44 (1.75)	0.82 (0.51)	81% (12%)	90% (8%)	95% (5%)
Phase II	8.01 (2.43)	0.39 (0.24)	79% (8%)	74% (15%)	88% (10.%)

* Reading time is in hours. Target ratio is non-monitored/monitored target frequency. Word M is word test accuracy scored by the experimenter in monitored sessions. Comp. NM is comprehension accuracy in non-monitored sessions. Comp. M is accuracy in monitored sessions. Standard deviations are in parentheses.

In summary, most subjects in Phase I accurately targeted misread words in both monitored and independent sessions. On the intermittent tests, their accuracy in reading previously targeted words and answering comprehension questions was quite high. Unfortunately, the average reading time on the system was far less than we hoped for because of a late start in the semester and frequent interference from competing school activities. Nevertheless, there were substantial and statistically significant training effects that are described in section 2 below.

The powerful training effects in Phase I (see below) led us to conclude that we could stretch our resources and effectively train a much larger sample for analyses of subtype by treatment interactions. Phase II included a total of 119 trained and 42 control children studied in either the fall 89 or spring 90 semesters. Eight of the trained subjects who had less than four hours of reading time were not included in the analyses. The mean reading time for the remaining 111 subjects was 8.1 hours with 14.2 total hours on the system, somewhat longer than in Phase I. However, because of limited personnel resources and a larger number of subjects in Phase II, there was about 40% less initial training for targeting difficult words and the follow up monitored sessions were about half as frequent. This lower level of pretraining and monitoring may have led to the significant declines from Phase I to Phase II in non-monitored/monitored target ratio ($F(1,134) = 41.79$, $p < 0.0001$) and accuracy in monitored ($F(1,134) = 4.59$, $p < 0.01$) and non-monitored ($F(1,134) = 33.1$, $p < 0.001$) comprehension tests (see Table 2). There were no significant differences between Phase I and II in monitored word recognition ($F(1,134) = 0.78$, $p > 0.05$). When the samples for the measures in Table 2 were combined across Phase I and Phase II, there were no significant differences in these variables across the three feedback conditions (all $p > 0.05$)

2. *Differences between Phase I and II nonword gains*

We attempted to balance the subjects' average nonword pre-test scores across the treatment conditions within Phase I and Phase II (see Table 3). The slight imbalance across the feedback conditions arose because of subject attrition. The control group in Phase I had somewhat higher pretest scores due to an error in balancing the groups. However, gain scores were not significantly correlated with pre-test scores. The correlation between pretest and gain scores across subjects in all conditions was $r = -0.04$.

Pre-test scores across all conditions in Phase II (mean = 40.3; sd = 19.2) averaged lower than in Phase I (mean = 46.9; sd = 18.9). This was associated with a slightly more severe average reading deficit and slightly younger mean age for subjects in Phase II, which included the Fall semester (see Table 1).

Table 3. Gains in nonword reading for feedback and control conditions in Phase I and Phase II*

Phase	Whole word	Syllable	Onset-rime	Control
Phase I				
Pre-test	44.3 (17.3)	40.1 (23.2)	48.7 (13.2)	56.2 (19.5)
Gain	6.0 (6.4)	13.7 (9.2)	17.5 (11.3)	2.8 (8.3)
N	8	10	9	8
Phase II				
Pre-test	39.9 (16.3)	37.0 (18.3)	42.1 (22.6)	42.9 (18.8)
Gain	9.9 (10.0)	10.7 (10.9)	7.6 (8.6)	2.5 (8.3)
N	33	36	42	38
Phase I and II				
Pre-test	40.7 (16.4)	37.7 (19.2)	43.2 (21.3)	45.2 (19.4)
Gain	9.1 (9.4)	11.3 (10.6)	9.4 (9.8)	2.5 (8.1)
N	41	46	51	46

* Gain scores are pre-test percent correct subtracted from post-test percent correct. Standard deviations are in parentheses.

Nonword gain scores were analyzed in a condition by phase ANOVA* The main effect of condition was highly significant due to the very small gain for the control group ($F(3,176) = 7.55$, $p < 0.001$; contrast of control vs trained conditions: $t(180) = 4.56$, $p < 0.001$). The main effect for phase showed a marginally significant decline from Phase I (9.61) to Phase II (7.75) ($F(1,176) = 2.64$, $p = 0.13$), in spite of the longer training time in Phase II. In addition, the interaction between phase and condition just reached significance ($F(3,176) = 2.64$, $p = 0.05$). Planned comparisons revealed that the whole-word condition was significantly better than the untrained control condition in Phase II ($t(176) = 3.33$, $p < 0.01$), although that had not been the case in Phase I ($t(176) = 0.65$, $p > 0.05$). Gains in the syllable condition were significantly greater than in the control condition in both Phase I and II ($t(176) = 3.2$ and 3.77 respectively, both $p < 0.01$).

Having established that the significant main effect of condition was due to the small gains in the control group, a second condition by phase ANOVA was performed without the control group to focus on effects related to feedback condition in training. Gains averaged across the three feedback conditions showed a marginally significant decline from Phase I to Phase II (12.6 vs. 9.3; $F(1,132) = 2.52$, $p = 0.11$). The main effect for feedback condition was not significant ($F(1,132) = 0.54$, $p > 0.05$). The most significant effect was the interaction between phase and feedback condition ($F(2,132) = 3.45$, $p < 0.05$). This interaction was primarily due to the onset-rime condition, which was the most beneficial feedback condition in Phase I but the least helpful feedback condition in Phase II; ($t(176) = 2.87$, $p < 0.01$, for Phase I onset-rime vs. Phase II onset-rime). Phase II onset-rime was still significantly better than the untrained control condition ($t(176) = 2.48$, $p < 0.05$).

In summary, we hypothesize that the overall decline in nonword training effects between Phases I and II, in spite of the longer training time in Phase II, was due to the substantial reduction in experimenter-monitored sessions and associated reduced training on targeting behavior. This change apparently led to the significant reduction in targeting ratio in Phase II. It is possible that the gains from onset-rime feedback are particularly sensitive to the amount of experimenter training and monitoring of targeting behavior, thus accounting for the sharp reduction in the onset-rime gain score for Phase II. However, an analysis of the effects of reading-deficit severity, presented later in the chapter, will suggest that onset-rime subjects in Phase I

---

* All of the ANOVAs on gain scores in the present paper were complemented by ANOVAs using the pre- and post-tests as repeated measures. The results of the repeated-measures analyses consistently revealed a significant main effect of test (e.g. subjects got better from pre- to post-test) and no significant differences in pre-test scores across the conditions. The F ratios and significance levels for the test by condition interactions were nearly the same as for the analyses of gain scores. Therefore, to simply the reporting of the critical gain scores and their interactions, we elected to report the ANOVAs for gain scores instead of the repeated-measures ANOVAs.

may have had greater gains because they were less severely reading disabled than the onset-rime subjects in Phase II. Finally, the absence of a significant training benefit to nonword reading from whole-word feedback (compared to the untrained control condition) in Phase I seems to have been due to chance, since this contrast was significant in Phase II, and in the combined sample.

3. *Differences between Phase I and II word-recognition gains*

To obtain a more reliable estimate of each subject's word recognition, the number of correct items up to the subject's final list position in the PIAT and in the Timed Word-Recognition tests were added together to provide a composite score for the present analyses. The general pattern of pre-test scores for word recognition was similar to that described above for nonword reading (see Table 4). The average pre-test score in Phase I (99.6; sd = 36.8) was higher than in Phase II (87.5; sd = 34.3), again reflecting the slightly younger age and more severe deficits in Phase II. Pre-test scores across the feedback conditions varied slightly, notably in the better pre-test performance of subjects in the control condition relative to the conditions which would receive training. However, pre-test scores were not significantly correlated with gain scores. The correlation between pre-test and gain scores for word recognition across all subjects was $r = 0.05$.

Table 4. Gains in word recognition for feedback and control conditions in Phase I and Phase II*

Phase	Whole word	Syllable	Onset-rime	Control
Phase I				
Pre-test	99.1 (45.4)	88.4 (39.1)	89.3 (34.0)	121.6 (22.3)
Gain	16.7 (8.1)	17.0 (10.8)	24.7 (7.3)	8.6 (6.9)
N	7	10	10	10
Phase II				
Pre-test	91.6 (36.1)	84.6 (30.5)	78.3 (35.4)	95.8 (33.6)
Gain	17.1 (10.2)	14.9 (11.6)	16.8 (10.8)	10.1 (9.7)
N	32	35	41	41
Phase I and II				
Pre-test	93.0 (37.5)	85.4 (32.1)	80.4 (35.1)	100.8 (33.2)
Gain	17.0 (9.8)	15.4 (11.3)	18.4 (10.6)	9.8 (9.2)
N	39	45	51	51

* Gain scores are the increase in number of words read correctly on the PIAT and Timed Word Recognition tests. Standard deviations in parentheses.

The subjects' word-recognition gain scores were analyzed in a phase by condition ANOVA. A slight decline in word recognition gains appeared from Phase I to Phase II despite longer reading times, but the effect of phase was not statistically significant ($F(1,178) = 1.37$, $p = 0.24$). The main effect of treatment condition was highly significant ($F(3,178) = 6.70$, $p < 0.001$), due mostly to the smaller gains for subjects in the control group compared to subjects in the trained groups ($t(182) = 4.19$, $p < 0.001$). Unlike the results described above for nonword gains, there was no significant interaction between phase and treatment condition ($F(3,178) = 1.31$, $p = 0.27$).

A second ANOVA was performed without the control group to focus on effects associated with feedback condition. The main effect of condition was no longer significant ($F(2,127) = 0.94$, $p = 0.40$), indicating nearly equivalent gains from whole-word, syllable, and onset-rime feedback. The effect of phase was marginally significant (Phase I = 19.6, Phase II = 16.2; $F(2,127) = 2.34$, $p = 0.13$). The interaction between phase and feedback condition was not significant ($F(2,127) = 1.07$, $p = 0.35$). As in the nonword gains, the largest and only significant difference between Phase I and II was in the onset-rime condition ($t(178) = 2.19$, $p < 0.05$).

Interpretation of the gains made by trained subjects on post-test word recognition requires knowing whether the words gained were targeted for feedback at some point in the training. Reading materials varied for all subjects based on ability and interest, and different words were studied by different children. Across all subjects, only 12% of the actual words gained on post-tests that had not been known on pretests, had actually been words targeted during the reading of stories. The fact that 88% of words gained on post-tests had *not* been targeted could mean that these words were seen and learned through reading or other instruction outside the computer training program. A more interesting possibility is that the feedback conditions helped to improve subjects' generalizable phonological coding skills (as in nonword reading) that could support their independent word recognition on the post-test. In support of this hypothesis, there was a significant correlation between gains in nonword reading and gains in word recognition across all subjects ($r = 0.44$, $p < 0.01$)

In summary, all three feedback conditions showed a significant advantage over the untrained control group's gains in word recognition, but there were no significant differences across the feedback conditions and there was no significant interaction between condition and deficit severity. The marginally significant decline in word-recognition gains between the Phase I and II trained conditions, in spite of the longer reading time in Phase II, mirrors the results for nonword gains. The results suggest that the level of training and monitoring of targeting behavior is important for improvement in both word recognition and phonological decoding skills. Finally, it appeared that at least part of the gains in word recognition from the feedback conditions may have been due to correlated increases in subjects' phonological decoding skills.

C. *Combined sample interaction with deficit severity*

There was a wide range of deficit severity in word recognition for subjects in Phase I and II. We now test the hypothesis that disabled readers' gains from the different feedback conditions may interact with their deficit severity. The Phase I and II samples were combined to provide a more powerful test.

1. *Nonword gains and deficit severity*

Reading-deficit severity was assessed by dividing each child's PIAT word-recognition grade-level by their grade level in school. Thus, a fourth grader who read words at the second grade level according to the PIAT national norms would have a deficit severity of 0.5. This procedure provides comparable estimates of deficit severity across the end of 2nd to 6th grade children in our sample. The mean level of deficit severity in the combined sample was 0.78. We noted in our earlier discussion of the subjects that this may not seem very severe, but it is when compared with the mean level of 1.2 for children in the Boulder schools, where socioeconomic and education levels are higher than average.

To assess the effects of deficit severity, subjects with a deficit severity below 0.7 were in the more severe group (mean = 0.56; sd = 0.10), while those at or above 0.7 were in the less severe group (mean = 0.94; sd = 0.16). Results for nonword gain scores across the different treatment conditions by deficit severity are presented in the top half of Table 5. An analysis of variance revealed a highly significant main effect of level of severity ($F(1,176) = 7.95$, $p < 0.001$). Excluding the control group which showed very little gain, the more severely disabled group averaged 7.8% and the less severely disabled group averaged 11.8% improvement in nonword reading. As in the previous analysis by phase, the main effect of treatment condition was also highly significant due to the very small gains for the control group ($F(3,176) = 9.46, p < 0.001$).

Most important, the interaction between level of severity and treatment

Table 5. Nonword and word recognition gain scores for feedback and control conditions at two levels of deficit severity in word recognition*

	Whole word	Syllable	Onset-rime	Control
Nonword				
Most severe	7.5 (9.5)	12.2 (13.3)	5.0 (7.7)	−0.5 (9.5)
N	17	19	28	11
Least severe	10.3 (9.4)	10.7 (8.3)	14.7 (9.5)	3.5 (7.5)
N	24	27	23	35
Word recognition				
Most severe	17.1 (9.4)	11.9 (11.1)	15.1 (10.3)	7.7 (8.2)
N	17	18	28	12
Least severe	16.9 (10.2)	17.7 (11.0)	22.3 (9.8)	10.5 (9.5)
N	22	27	23	39

* Nonword gain scores are % correct pre-test subtracted from % correct post test. Word-recognition gain-scores are increase in words correct. Standard deviations in parentheses.

condition was significant ($F(3,176) = 2.98$, $p < 0.05$). (This pattern was present in both phases of the study, but was not statistically significant in the phases by themselves due to the reduced sample size.) It is clear from Table 5 that most of the significant interaction for the combined sample is due to the large difference in gain scores associated with deficit severity in the onset-rime feedback condition (5.0 vs. 14.7; $t(176) = -3.76$, $p < 0.001$). Within the more severely disabled group, onset-rime feedback yielded the smallest gain of the three feedback conditions, significantly less than the syllable group ($t(176) = 2.62$, < 0.05), and not significantly greater than the control group. The 63% advantage of the syllable group over the whole-word group was only marginally significant ($t(178) = 1.54$, $p = 0.13$). In contrast, within the *less* severely disabled group, onset-rime feedback yielded the largest gain of the three feedback conditions, although the 39% greater gain from onset-rime compared to whole-word feedback was only marginally significant ($t(176) = 1.83$, $p = 0.07$).

The above results suggest a possible explanation of the large difference between Phase I and Phase II gain scores in the onset-rime condition. Only 2 of the 9 Phase I onset-rime subjects were in the more severe group, and they had an average gain score of only 7%. The other 7 subjects in the less severe group had an average gain score of 20.5%! The distribution of Phase II onset-rime subjects across level of severity was in the opposite direction. Twenty six subjects were in the more severe group with a mean gain score of 4.8% and 16 subjects were in the less severe group with mean gain scores of 12.2%. The preponderance of more severely disabled onset-rime subjects in Phase II probably was a major factor contributing to their lower overall gains compared to the Phase I subjects (see Table 3).

In summary, the general advantage of the feedback conditions over the control condition was qualified by a significant interaction between condition and deficit severity. The nonword gain from onset-rime segmentation in the more severe group was relatively small and not significantly different from the control group, but it was the best condition for the less severely disabled subjects. In contrast, syllable segmentation was the best condition for the more severely disabled subjects.

2. *Word-recognition gains and deficit severity*

Word-recognition gains were analyzed in a treatment-condition by deficit-severity ANOVA (see lower half of Table 5). As in the earlier analysis by phase, the main effect of treatment condition was highly significant ($F(3,178) = 8.50$, $p < 0.001$), due to the substantially smaller gain of the control group ($t(182) = 4.19$, $p < 0.0001$). (Treatment condition was not significant when the control group was excluded from the analysis ($F(2,129) = 1.45$, $p > 0.05$.) The main effect of deficit severity was also significant (more severe = 13.3 (sd 10.2), less severe = 15.9 (sd 10.8); $F(1,178) = 7.26$, $p < 0.001$). However, the interaction between severity and treatment condition that was significant for nonword gains was *not* significant for gains in word recognition ($F(3,178) = 1.13$, $p = 0.336$). Moreover, in the more severe group, the pattern for word-recognition gain scores was different from that seen for nonword reading: In the more severe group, onset-rime feedback was nonsignificantly better than syllable feedback ($t(178) = -1.07$, $p > 0.05$), and whole-word feedback was non-significantly better than both syllable ($t(178) = 1.54$, $p = 0.13$) and onset-rime feedback ($t(178) = 0.64$, $p = 0.52$). In the less severe group, the pattern for word recognition was more similar to that found for nonword gains: Onset-rime feedback was non-significantly better than syllable ($t(178) = 1.58$, $p = 0.12$) or whole-word feedback ($t(178) = 1.78$, $p = 0.08$).

3. *Daily word-recognition and comprehension performance*

A different view of word recognition can be obtained from an analysis of the number of targeted words read correctly in the intermittent word-check tests. Across both phases, the average percent correct for targeted words that were initially misread in the monitored tests was 78.0%, 78.5%, and 80.3% respectively in the whole-word, syllable, and onset-rime conditions ($F(2,129) = 0.777$, $p > 0.05$). Thus, there was substantial and nearly equivalent short-term learning of targeted words across the three feedback conditions. In addition, the average percent correct comprehension scores in the whole-word (79.7%), syllable (75.9%), and onset-rime (75.2%) conditions were not significantly different ($F(2,129) = 0.802$, $p > 0.05$).

4. *PIAT reading comprehension and spelling*

Treatment effects for PIAT spelling and comprehension were also analyzed in a condition by deficit-severity ANOVA. Grade-equivalent gain scores for PIAT Spelling (whole-word = 0.07, syllable = 0.30, onset-rime = 0.16, and control = −0.19) were significantly different, largely due to the *decline* in performance of the control group ($F(3,161) = 3.49$, $p < 0.05$). The unexplained decline of the control group along with the poor published reliability of this test (0.65, Dunn & Markwardt 1970) lead us to have little confidence in these results. Effects due to deficit severity ($F(1,161) = 0.23$, $p > 0.05$), and the interaction between deficit severity and condition ($F(3,161) = 0.67$, $p > 0.05$), were not significant.

Grade-equivalent gain scores for PIAT Reading Comprehension in the whole-word (0.73), syllable (0.75), onset-rime (0.71), and control (0.49) conditions were not significantly different ($F(3,163) = 0.63$, $p > 0.05$), nor was the main effect of deficit severity or its interaction with condition ($p > 0.05$). These null results were not surprising in view of the low reliability of 0.64 reported for this test (Dunn & Markwardt 1970).

D. *Pig-latin and phoneme-deletion gains and correlations with other variables*

The pig-latin and phoneme-deletion tasks were included in the test battery to see if orthographic and speech feedback would improve subjects' phonological awareness. In addition, we wondered if subjects' initial level of phonological awareness would be related to gains in phonological decoding and word recognition. Subjects in Phase I and the fall semester of Phase II were given the pig-latin task. Subjects in both semesters of Phase II were given the phoneme-deletion task. We will first present results from the pig-latin task.

1. *Pig latin*

The gain scores for pig latin were analyzed in a condition by deficit-severity ANOVA. Average gain scores were not significantly different across the conditions (whole-word = 18.5%, syllable = 13.8%, onset-rime = 18.0%, and control = 7.0%; $F(3,100) = 1.97$, $p = 0.122$). The relation between deficit-severity and pig-latin gains was significant (more severe = 10.2%, less severe = 19.2%; $F(1,100) = 6.68$, $p < 0.05$). The interaction between deficit severity and condition was not significant ($F(3,100) = 0.598$).

Although subjects' gain scores in pig latin were not significantly different by condition, their pig-latin pre-test and gain scores were significantly related to pretest and gain scores for phonological decoding and word recognition. The age-adjusted second-order correlations between pig-latin pre-test, pig-latin gains, nonword pre-test, nonword gains, word-recognition pre-test, and word-recognition gains are presented in Table 6. Correlations in the top row for each variable are for the 111 subjects across all conditions who had complete data for all of the variables. Correlations in the bottom row for each variable are for the 30 subjects trained in the onset-rime condition, which showed a uniquely strong correlation between pig-latin pretest and gain scores for nonwords. Corresponding correlations between pig-latin pre-test and nonword gains were $r = 0.00$ for whole-word, $r = 0.37$ for syllable, and $r = 0.16$ for the control group.

Not shown in Table 6 are the age-adjusted partial correlations between pig-latin pre- and post-tests ($r = 0.67$), nonword pre- and post-tests ($r = 0.89$), and word-recognition pre- and post-tests ($r = 0.94$). These correlations provide low-boundary estimates for long-term test reliability within the restricted range of the present sample. The estimates would have been somewhat reduced by any differential treatment effects that influenced post-test scores.

The small but significant negative correlations between pig-latin pretest and pig-latin gain scores may be due to regression to the mean for subjects who initially scored unusually low on the pre-test, perhaps because they did not sufficiently understand the task. The pig-latin pre-test correlations were slightly higher with the nonword than the word-recognition pre-tests, a pattern that was also observed with the phoneme-deletion task discussed below. The most interesting results in Table 6 are the significant correlations between subjects' pig-latin pre-test scores and their gains in nonword reading, particularly in the onset-rime feedback condition.

2. *Phoneme deletion*

The gain scores for phoneme deletion were analyzed in a condition by deficit-severity ANOVA. Average gain scores were not significantly different

Table 6. Correlations between pig latin and other variables

	Pig-latin gain	Nonword pre-test	Nonword gain	Word rec. pre-test	Word rec. gain
Pig-latin pre-test	−0.16*	0.37**	0.32**	0.29**	0.15
	−0.31*	0.33*	0.55**	0.12	0.39*
Pig-latin gain		0.13	0.23**	0.10	0.26**
		0.15	0.20	0.12	0.30*
Nonword pre-test			0.11	0.71**	0.29**
			0.27	0.69**	0.39**
Nonword gain				0.13	0.50**
				0.29	0.68**
Word rec. pre-test					0.13
					0.24

Correlations in the top row for each variable are for the 111 subjects across all conditions who had complete data for all of the variables. Correlations in the bottom row are for the 30 subjects trained in the onset-rime condition. * = $p < 0.05$; ** = $p < 0.01$.

across the conditions (whole-word = 5%, syllable = 12%, onset-rime = 8%, and control = 5%; $F(3,110) = 1.24$, $p > 0.05$). The relation between deficit-severity and phoneme-deletion gains was not significant (more severe = 8%, less severe = 7%; $F(1,100) = 0.012$, $p > 0.05$). The interaction between deficit severity and condition also was not significant ($F(3,100) = 0.874$, $p > 0.05$).

Although subjects' gain scores in phoneme deletion were not significantly different by condition, their phoneme-deletion pre-test and gain scores were significantly related to pre-test and gain scores for nonwords and word recognition, showing a similar pattern of correlations to that reported above for pig latin. The age-adjusted second-order correlations between phoneme-deletion pre-test, phoneme-deletion gains, nonword pre-test, nonword gains, word-recognition pre-test, and word-recognition gains are presented in Table 7. Correlations in the top row for each variable are for the 110 subjects across all conditions who had complete data for all of the variables. Correlations in the bottom row for each variable are for the 28 subjects trained in the onset-rime condition, which showed a uniquely strong correlation between phoneme-deletion pre-test and gain scores for nonwords. Corresponding correlations with nonword gains were $r = 0.30$ for the whole-word group, $r = 0.20$ for the syllable group, and $r = 0.15$ for the control group.

Not shown in Table 7 are the age-adjusted partial correlations for the entire sample (N = 110) between phoneme-deletion pre-test and post-test ($r = 0.78$), nonword pre- and post-tests ($r = 0.89$), and word-recognition pre- and post-tests ($r = 0.93$). These correlations provide low-boundary estimates for long-term test reliability within the restricted range of the present sample, reduced by any differential treatment effects.

Table 7. Correlations between phoneme deletion and other variables

	Phoneme gain	Nonword pre-test	Nonword gain	Word rec. pre-test	Word rec. gain
Phoneme	−0.13	0.58**	0.27**	0.42**	0.32**
Pre-test	−0.12	0.68**	0.47**	0.46**	0.71**
Phoneme		0.11	0.24**	−0.01	0.02
Gain		0.26	0.34*	0.25	−0.12
Nonword			0.06	0.69**	0.10
Pre-test			0.34*	0.75**	0.42*
Nonword				0.10	0.48**
Gain				0.26	0.55**
Word rec.					−0.03
Pre-test					0.17

Correlations in the top row for each variable are for the 110 subjects across all conditions who had complete data for all of the variables. Correlations in the bottom row are for the 28 subjects trained in the onset-rime condition. * = $p < 0.05$; ** = $p < 0.01$.

The phoneme-deletion pre-test correlations were higher with the nonword than the word-recognition pre-tests, a pattern that was also observed with the pig-latin task discussed earlier. Although the differences between correlations for the present samples are not statistically significant, the pattern of results was consistent with that found in our analysis of twin data (Conners & Olson 1991). The consistently high correlations with nonword reading suggest that subjects' phonological awareness plays a particularly important role in the development of their phonological decoding skills.

The most interesting results in Table 7 are the significant correlations between subjects' phoneme-deletion pre-test scores and their gains in nonword reading and word recognition. Following the pattern found for pig latin, the correlations were highest within the onset-rime feedback condition. Also, phoneme-deletion gains were significantly correlated with gains in nonword reading across the whole sample and more strongly within the onset-rime group. The absence of significant correlations between gain scores in phoneme deletion and word recognition emphasizes the unique relation between the development of phonological awareness and phonological decoding.

3. *Independent prediction of nonword and word-recognition gains*

A final series of partial correlations was performed to see if the predictive relations of pig-latin and phoneme deletion to nonword and word-recognition gains was due to their association with other variables. Third-order correlations were calculated after partialing age with IQ, age with the continuous measure of deficit severity on the PIAT word-recognition pre-test, and age with the combined word-recognition pre-test. Even though the phoneme-deletion and pig-latin pre-tests were significantly correlated with these three variables, the third-order partial correlations with nonword and word-recognition gains differed by no more than 0.05 from those partialed on age only in Table 6 and 7. A final partial correlation controlled for subjects' age and nonword pre-test scores. Even though the nonword pre-test was highly correlated with the phonological awareness pre-tests, the partial correlations between phonological awareness and the difference scores were not significantly different. It is clear from these results that subjects' phonological-awareness skills were unique predictors of their rate of improvement during training, particularly in the onset-rime condition.

E. *Interactions between phonological awareness and condition*

Previous analyses of nonword and word-recognition gain scores revealed a significant interaction between condition and deficit severity in word recognition (Table 5). However, the pattern of correlations reported in the previous section suggests that gains in the different conditions might also interact with subjects' level of phonological awareness reflected in our pig-latin and phoneme-deletion tasks.

1. *Pig-latin*

Subjects in Phase I and the fall semester of Phase II who had been tested on pig-latin were divided into a group above and a group below the mean of their age-adjusted pig-latin pre-test score. More subjects were in the lower group because of floor effects on the pig-latin test. Condition by level of pig-latin ANOVAs were performed for nonword and word-recognition gain scores.

(a) *Nonword gains.* The means for nonword gains are presented in the upper half of Table 8. There was a significant effect of condition ($F(3,119) = 3.23$, $p < 0.05$). The effect of pig-latin level (high pig-latin = 6.9%, low pig-latin = 12.1%) was also significant ($F(1,119) = 7.28$, $p < 0.01$). The interaction between condition and pig-latin level was only marginally significant ($F(3,119) = 2.25$, $p = 0.086$). This marginal effect was due to the greater gains for both segmented feedback conditions in the high pig-latin group.

(b) *Word-recognition gains.* The means for word-recognition gains are presented in the lower half of Table 8. The effect of condition was significant ($F(3,118) = 4.17$, $p < 0.01$). The effect of pig-latin level (high pig-latin = 13.9, low pig-latin = 18.7) was significant ($F(1,118) = 4.92$, $p < 0.05$). The interaction between condition and pig-latin level was also significant ($F(3,118) = 3.16$, $p < 0.05$). Similar to the pattern for gains in nonword reading, there were substantially increased gains in word recognition for the syllable ($t(118) = 2.02$, $p < 0.05$) and onset-rime ($t(118) = 3.03$, $p < 0.01$) subjects in the higher pig-latin level compared to the lower level, while

Table 8. Nonword and word-recognition gain scores for feedback and control conditions at two levels of pig latin*

	Whole word	Syllable	Onset-rime	Control
Nonword				
Low pig latin	9.1 (9.0)	8.6 (10.8)	5.6 (8.9)	3.1 (9.7)
N	21	23	22	15
High pig latin	8.6 (11.9)	15.1 (10.7)	15.9 (10.1)	1.8 (8.9)
N	9	13	17	7
Word recognition				
Low pig latin	18.7 (11.8)	12.9 (11.9)	14.2 (9.6)	9.5 (7.5)
N	19	22	21	16
High pig latin	14.2 (8.1)	20.3 (9.7)	24.5 (10.0)	8.4 (15.2)
N	9	13	18	8

* Nonword gain scores are % correct pre-test subtracted from % correct post test. Word Recognition gain scores are increase in correct words. Standard deviations in parentheses.

subjects in the whole-word and control conditions showed slightly *smaller* gains in the higher pig-latin level compared to the lower level.

2. *Phoneme deletion*

Subjects in Phase II were divided into groups above and below the mean of their age-adjusted phoneme-deletion score. Condition by level of phoneme-deletion ANOVAs were performed on their nonword and word-recognition gain scores. Means for the different conditions at high and low levels of phoneme deletion are presented in Table 9.

(a) *Nonword gains.* The main effect of condition was significant ($F(3,111) = 5.82$, $p < 0.01$). The effect of phoneme-deletion level (low phoneme-deletion = 5.1, high phoneme-deletion = 10.5) was also significant ($F(1,111) = 10.36$, $p < 0.01$). However, the interaction between condition and phoneme-deletion level did not approach significance ($F(3,111) = 0.58$, $p > 0.05$).

(b) *Word-recognition gains.* The main effect for condition was significant ($F(3,111) = 3.33$, $p < 0.05$). The main effect of phoneme-deletion level was also significant (low = 10.6, high = 16.9; $F(1,111) = 9.28$, $p < 0.01$). Again, the interaction between condition and level of phoneme-deletion did not approach significance ($F(1,111) = 0.96$, $p > 0.05$).

3. *Summary of results of phonological awareness*

(a) *Interactions.* Level of phoneme deletion did not interact with condition for gains in word recognition or nonwords. All conditions showed greater gains for the higher level of phoneme deletion. This result is in contrast to

Table 9. Nonword and word-recognition gain scores for feedback and control conditions at two levels of phoneme deletion*

	Whole word	Syllable	Onset-rime	Control
Nonword				
Low phoneme	5.7 (10.7)	9.2 (12.2)	5.0 (9.3)	0.0 (9.3)
N	9	20	15	17
High phoneme	15.3 (4.4)	12.8 (9.6)	8.9 (6.9)	5.6 (5.9)
N	14	12	18	14
Word recognition				
Low phoneme	16.4 (10.4)	12.2 (9.2)	9.8 (8.1)	6.7 (9.9)
N	9	20	15	18
High phoneme	18.5 (10.6)	16.2 (13.9)	20.4 (11.3)	11.5 (7.4)
N	14	11	18	14

* Nonword gain scores are % correct pre-test subtracted from % correct post test. Word recognition gain scores are increase in correct words. Standard deviations in paretheses.

the significant interaction found in the analyses by level of pig-latin, wherein only the segmented feedback conditions showed substantially greater gains at the higher level of pig-latin. Therefore, we must question the validity of the pig-latin interaction. For the 72 subjects in the fall semester of Phase II who had both the phoneme-deletion and pig-latin tests, the correlation between these two tests was $r = 0.42$, which is respectable in view of the tests' modest reliabilities discussed earlier. We would thus expect similar interactions with condition for level of pig latin and level of phoneme deletion. Replication would be required to have confidence in the interaction found for pig-latin.

(b) *Main effects and correlations.* Level of phoneme deletion and pig-latin were both associated with strong differences in gain scores for nonwords and word recognition averaged across the conditions. These differences were generally stronger than those reported in previous analyses with level of deficit severity in PIAT word-recognition. The same result is expressed in the significant correlations for nonword and word-recognition gains with pig-latin and phoneme-deletion pre-test scores (Tables 6 and 7). Moreover, these correlations were essentially unchanged when partialed on subjects' IQ scores, word-recognition pre-tests, or nonword pre-test.

V. DISCUSSION AND CONCLUSION

A. *General training effects*

We hypothesized that reading on the computer with speech feedback would significantly improve disabled readers' phonological decoding and word recognition. The results of our first two long-term studies provided strong support for this hypothesis. From the beginning to the end of a semester, control subjects who received their normal course of language-arts instruction exhibited very small gains in their phonological decoding of nonwords. The subjects who displaced part of their language-arts instruction with computer training exhibited gains in a test of phonological decoding that were about four times larger than the control group.

Reading on the computer also resulted in significantly greater gains in word recognition. The trained subjects' gains in number of items read correctly on the post-tests averaged nearly twice those of the control group. Grade-level gains on the standardized PIAT word-recognition test averaged 0.33 for the control subjects and 0.64 for the computer-trained subjects across the semester.

The analysis of differences in gains between Phase I and II highlighted the importance of adequate pretraining, monitoring, and encouragement by the experimenter. On average, the best gains were obtained from subjects who got the most individual attention and who started with relatively higher phonological awareness. This seemed to be particularly important for the segmented feedback conditions. It was clear that the computer can provide effective assistance when children are reading independently, but it does not eliminate the need for a teacher to guide and encourage the child in the use of the system.

B. *Limitations in detecting gain-score differences across feedback conditions*

Before discussing the differences in training effects across feedback conditions and interactions with subjects' pre-test scores, we need to point out some limitations of the present studies. The general problem of statistical power for the analyses of gain-score differences is apparent from their large standard deviations. There were three general factors that contributed to this variance. The first was the reliability of the pre- and post-test measures. For example, the correlation between the first and second blocks of nonwords given on successive days in the pretest was $r = 0.88$. This suggests a respectable level of reliability within the constrained ability-range of our sample, but the standard deviation of subjects' difference in percent correct between the first and second blocks was 10.3%! Thus, it is apparent that test-error variance was substantial in relation to effect size, even though the subjects' nonword score was the average of the two blocks.

Larger effects of training would likely have led to greater power for comparisons between conditions. The effects of training were quite respectable considering the subjects' average of only eight hours reading on the computer. Training time was limited by the half hour per day the schools would allow for a pull-out research study. Teachers reasonably did not want the computer-trained subjects to completely miss practice in spelling and writing during the semester. Additional constraints on training time included time required for getting children referred for the study, obtaining parental permission, pre-testing, pre-training, and post-testing in the span of a semester. Also, there were a large number of missed sessions due to holidays, in-service days, and school trips. We had contemplated training the subjects over a full year, but then the sample size for trained subjects would have been compromised because of our limited equipment and personnel resources. We also felt uncomfortable that the untrained control subjects would not have access to the computers all year, after we have identified them as having needs and after their parents had given permission to participate in the study. We are currently exploring the use of the system in the homes of reading-disabled twins from our behavior-genetic studies. Much longer training times will be possible in the home study and this should improve our ability to detect differences in training effects across the different feedback conditions.

The third factor leading to the large variance within conditions was the real variability in subjects' response to the training. Motivation seemed to be a problem for a minority of subjects. Others were highly motivated. Also, subjects varied in the amount and type of remedial instruction outside of their computer training. We found it difficult to categorize and quantify the instructional context to assess its effects on learning in the training conditions. Perhaps related to the instructional context as well as constitutional variables, we informally observed that subjects varied in the way they used the feedback within each condition. For example, some subjects in the whole-word condition were clearly and often successfully attempting to sound out difficult words. Others did not seem to apply their limited decoding skills in this condition. In the segmented feedback conditions, subjects also varied in their apparent attention to the segments. Some would use the initial segment in feedback along with sentence context to quickly identify the word and move on. Others seemed to pay more attention to all of the segments before continuing their reading. These factors were difficult to quantify, but we were able to assess the strong effects of individual differences in deficit severity and phonological awareness that will be discussed later.

C. *Main effects of different levels of feedback segmentation*

The effects of different levels of feedback segmentation were a major focus of the research. Whole-word feedback provided information for word identification. Segmented feedback also supported the accurate identification of targeted words, but it provided additional explicit information about the relations between sub-word letter patterns and their corresponding speech sounds. We hypothesized that disabled readers would need the additional information from segmented feedback to improve their phonological decoding significantly. This appeared to be the case in our first small study (Phase I; Wise et al. 1989), wherein whole-word feedback was significantly worse than segmented feedback and not significantly better than the control condition in improving nonword reading. However, the results of a second larger study (Phase II) indicated significant and approximately equal gains for all feedback conditions. This result was qualified by the interactions with deficit severity and phonological awareness discussed below, but it was clear that whole-word feedback did result in significant improvement for most disabled readers' phonological decoding.

The significant gains in phonological decoding from whole-word feedback are consistent with current connectionist simulation-models which have demonstrated that the correspondence between sub-word letter patterns and their speech sounds can be implicitly learned from whole-word feedback (Seidenberg & McClelland 1989). These authors speculate that a similar developmental process occurs automatically in children without explicit attention to word segments. There is much evidence that children do learn to read new words by implicit analogy to parts of known words (Marsh et al. 1981). However, there may be significant additional benefits for disabled readers' phonological decoding if the type of segmented feedback is appropriately matched to the severity of their reading deficit and/or their level of phonological awareness.

D. *Interactions with deficit severity and phonological awareness*

We hypothesized that disabled readers' gains in phonological decoding and word recognition might interact with the severity of their reading deficits and with their level of phonological awareness. Comparable data on deficit severity in word recognition were available for all subjects in Phase I and Phase II, allowing a powerful analysis of effects for the combined sample. Small samples were used to assess gain-score differences associated with pig-latin, which was given only in Phase I and the fall semester of Phase II, and with phoneme deletion, which was given only in Phase II. These differences in samples may account for some of the differences discussed below between interactions observed with deficit severity, pig-latin, and phoneme deletion.

For deficit severity, the significant interaction with type of feedback for gains in phonological decoding was due to the following pattern: Onset-rime feedback resulted in the greatest gains for phonological decoding in the less severe group, and the smallest gains in the more severe group. In contrast, syllable feedback led to modest gains, similar to whole-word feedback, in the less severe group, but syllable feedback yielded the best gains in the more severe group.

The above pattern of results for onset-rime feedback may be due to the fact that the more severely disabled readers also tended to have lower levels of phonological awareness for the onset and rime segments in speech that were required for successful performance in the pig-latin task. This deficit may have limited their ability to learn the relations between orthography and speech for these segments in the onset-rime feedback condition. In contrast, the better phonological awareness of the less severely disabled readers may have provided a sufficient linguistic foundation for their substantially greater benefit from onset-rime feedback.

The results for onset-rime feedback in the low and high severity groups were generally consistent with results from the groups divided by low and high phonological-awareness. Onset-rime was also the worst training condition in both the low pig-latin and low phoneme-deletion groups, although differences in nonword gains across the feedback conditions were not statistically significant in these low groups. For the high pig-latin group, onset-rime feedback again yielded the best gains for both nonwords and word recognition, although its advantage over syllable feedback was not significant. For high phoneme-deletion, onset-rime feedback was slightly better than the other conditions for gains in word recognition, but it yielded nonsignificantly lower gains for nonwords than either whole-word or syllable feedback. This discrepancy may simply be due to problems with error variance in the gain scores to be discussed later, or it may be due to sample differences. Because phoneme deletion was given only to subjects in Phase II, the sample did not include subjects in Phase I who received more experimenter training and demonstrated substantial gains in nonword reading. For less severe disabled readers with high phonological awareness, the superiority of onset-rime feedback for gains in nonword reading may depend on adequate pretraining and monitoring by the experimenter. This question requires further study. However, it seems relatively clear that for more severe disabled readers with low phonological awareness, onset-rime feedback is probably not appropriate.

Syllable feedback was the best condition for the most severely disabled readers, resulting in phonological decoding gains 140% greater than from onset-rime feedback and 61% greater than whole-word feedback. Even severely disabled readers have been found to have adequate phonological awareness for syllable segments in speech (Rozin & Gleitman 1977). The severe group's adequate phonological awareness of the larger syllable segments may have provided a foundation for their greater gains in phonological decoding from feedback that highlighted the relations between the syllables' orthography and speech. Syllable feedback was also nonsignificantly better than the other conditions in the low phoneme-deletion group, but it yielded nearly the same level of nonword gains as whole-word feedback in the low pig-latin group. These inconsistent results demonstrate the need for replication of the above interactions.

For word-recognition gains, we might expect to see an interaction between deficit severity and feedback condition similar to that found for phonological decoding, but the interaction for gains in word recognition did not approach significance. Level of pig-latin interacted similarly with condition for gains in phonological decoding and word recognition, but this interaction was not replicated for level of phoneme deletion. It is hard to know what to make of these mixed results. The correlation between gains in phonological decoding and word recognition was only modest ($r = 0.46$) across all subjects. This correlation was diminished by the limited reliability for gain scores discussed below. Also, it is not clear that gains in phonological decoding in our nonword reading task would immediately influence the subjects' performance in our word-recognition tasks. Rapid responding was encouraged and we did not explicitly ask subjects to sound out unknown words. It may take a longer period of training and consolidation of phonological decoding skills before gains in these skills become closely linked to gains in rapid word recognition.

E. *Phonological awareness predicts gains in phonological decoding and word recognition*

Overall main-effects of disabled readers' pre-test performance on the phonological-awareness tasks indicated a strong association with gains in phonological decoding and word recognition. Disabled readers who were above the means on the phonological-awareness measures averaged about twice the gains in phonological decoding compared to disabled readers below the means. In addition, the subjects above the mean on phoneme deletion demonstrated substantially greater gains in word recognition. These results, from dividing disabled readers into high and low phonological-awareness groups, were also reflected in correlations with the continuous variables. Pig-latin and phoneme deletion exhibited the highest pre-test correlations with subjects' gain scores in phonological decoding and word recognition. The correlations were maintained when separately partialed on subjects' IQ, word-recognition, and phonological-decoding pre-test scores.

Other studies have shown that pre-schoolers' phonological awareness predicts their later success in reading, and that across the normal range, phonological awareness is correlated with reading ability (for reviews, see Wagner & Torgesen 1987; Wise & Olson 1991). The present study adds the information that *within* groups of disabled readers, benefits from a program of remedial instruction may be strongly influenced by level of phonological awareness at the beginning of training. Moreover, this influence is at least partly independent from the severity of disabled readers' deficits in phonological decoding and word recognition. Our results are consistent with the relatively poor response to training for an extreme phonological-deficit subgroup of disabled readers studied by Lyon (1985).

Research with identical and fraternal twins has revealed that disabled readers' group-deficit in phonological awareness is greater than expected from their reading level and it has a significant genetic etiology (Olson et al. 1989). Nevertheless, several studies have shown that direct training may significantly improve childrens' phonological awareness with resulting benefits in reading (Ball & Blachman 1988; Bradley & Bryant 1983; Lundberg, Frost & Petersen 1988; Alexander et al. 1991). We are currently exploring the use of computer speech for the direct training of disabled readers' phonological awareness prior to or in conjunction with reading on the computer.

We have designed a computer program for this purpose that incorporates the manipulation of phonemes within syllables as seen in the training program developed by Rosner (1971) and in part of the Auditory Discrimination in Depth (ADD) program developed by Lindamood & Lindamood (1969). (Lindamood is developing a laser video-disc system to help with the important articulatory-motor training unique to her program, and we hope to collaborate with her to incorporate this system for pre-training in phonemic awareness for half of our future subjects.) The spelling program described by Wise and Olson (this volume) may also be combined with the reading program to improve disabled readers phonological awareness, decoding, and spelling.

F. *Conclusion*

With few exceptions, the children in the school studies have been enthusiastic about reading on the computer. Teachers and parents report marked increases in the children's confidence and interest in reading that coincide with their real gains in phonological decoding and word recognition. We initially had to work hard to get our research program into the schools. Now there are more requests for the program than can be accommodated. However, there are important research questions that remain to be answered to achieve optimal benefits for each child. We need to replicate the patterns of differential gains across the feedback conditions related to deficit severity and phonological awareness, and we are exploring the use of computer speech for the remediation of disabled readers' deficits in phonological awareness. Further research is needed to determine the best ways to use talking computers with different children, but it is clear from the present studies and others reported in this volume that computer speech will play an increasingly important role in the remediation of reading disabilities.

ACKNOWLEDGEMENTS

We thank Mike Anstett for programming the systems, and Laura Kriho, Bonnie Houkal, Lauralyn Andrews, Maureen Terjak, Denise Arthur, Vicki Schneider, Beverly Budwit, Julie Kostuch, Jerry Ring, Jeanie McCray, Cathryn Burns, and Laurel Toeppen for training children in the schools to read on the computer during various phases of this project. This research was made possible by the support of teachers, administrators, children, and parents in the Boulder Valley School District. The research was financed through NICHD grant Nos. HD11683, HD22223, and HD27802. Portions of this paper were presented at the 1990 AERA meeting in Boston. Please send correspondence to Richard Olson or Barbara Wise, Department of Psychology, Campus box 345, University of Colorado, Boulder, CO 80309-0345, USA.

REFERENCES

Alexander, A. W., Anderson, H. G., Heilman, P. C., Voeller, K. S. & Torgesen, J. K. (1991). Phonological awareness training and remediation of analytic decoding deficits in a group of severe dyslexics, *Annals of Dyslexia* 41: 193—206.

Ball, E. W. & Blachman, B. A. (1988). Phoneme segmentation training: Effect on reading readiness, *Annals of Dyslexia* 38: 208—225.

Barron, R. W., Golden, J. O., Seldon, D. W., Tait, C. F. & Marmurek, H. C. (1990). Training prereading skills with DECtalk: Evidence for a bidirectional causal relationship between phonological awareness and literacy. Paper presented at the American Educational Research Association meetings, 16—20 April 1990.

Besner, D., Twilley, L., McCann, R. S. & Seergobin, K., (1990). On the association between connectionism and data: Are a few words necessary?, *Psychological Review* 97: 432—446.

Boder, E. & Jarrico, S. (1982). *The Boder test of reading-spelling patterns: A diagnostic screening test for subtypes of reading disability.* New York: Grune & Stratton.

Bradley, L. & Bryant, P. (1983). Categorizing sounds and learning to read, *Nature* 271: 746—747.

Campbell, R. & Butterworth, B. (1985). Phonologocal dyslexia and dysgraphia in a highly literat subject: A developmental case with associated deficits of phonemic processing and awareness, *Quarterly Journal of Experimental Psychology* 37: 435—475.

Conners, F. & Olson, R. K. (1990). Reading comprehension in dyslexic and normal readers: A component skills analysis, pp. 557—579, in: D. A. Balota, G. B. Flores d'Arcais & K. Rayner (eds.), *Comprehension processes in reading.* Hillsdale, NJ: Erlbaum.

DeFries, J. C., Olson, R. K., Pennington, B. F. & Smith, S. D. (1991). Colorado reading project: An update, pp. 58—87, in: D. Duane & D. Gray (eds.), *The reading brain: The biological basis of dyslexia.* Parkton, MD: York Press.

Dole, J. A., Duffy, G. G., Roehler, L. R. & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction, *Review of Educational Research* 61 (2): 239—264.

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

Fox, B. & Routh, D. (1980). Phonemic analysis and severe reading disability in children, *Journal of Psycholinguistic Research* 9: 115—119.

Goodman, K. (1967). Reading: A psycholinguistic guessing game, *Journal of the Reading Specialist* 6: 126—136.

Jorm, A. F. & Share, D. L. (1983). Phonological recording and reading acquisition, *Applied Psycholinguistics* 4: 103—147.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967). Perception of the speech code, *Psychological Review* 74: 431—461.

Liberman, I. Y. (1983). A language oriented view of reading and its disabilities, in: H. Myklebust (ed.), *Progress in learning disabilities,* Vol. 5. New York: Grune & Stratton.

Lindamood, C. H. & Lindamood, P. C. (1969). *Auditory discrimination in depth.* Boston, MA: Teaching Resources.

Lindamood, C. H. & Lindamood, P. C. (1979). *Lindamood auditory conceptualization test.* Hingham, MA: Teaching Resources Corporation.

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.

Lyon, G. R. (1985). Educational validation studies, pp. 228—253, in: B. P. Rourke (ed.), *Neuropsychology of Learning Disabilities* New York: Guslford Publications.

Marsh, G., Friedman, M., Welch, V. & Desberg, P. (1981). A cognitive-developmental theory of reading acquisition, in: G. E. MacKinnon and T. G. Waller (eds.), *Reading research: Advances in theory and practice*. Vol. 3. New York: Academic Press.

Olson, R. K. (1985). Disabled reading processes and cognitive profiles, pp. 215—244, in: D.

Gray and J. Kavanagh (eds.), *Biobehavioral Measures of Dyslexia.* Parkton, MD: York Press.

Olson, R., Foltz, G. & Wise, B. (1986). Reading instruction and remediation with the aid of computer speech, *Behavior Research Methods, Instruments, and Computers* 18: 93—99.

Olson, R. K., Gillis, J. J., Rack, J. P., DeFries, J. C. & Fulker, D. W. (1991). Confirmatory factor analysis of word recognition and process measures in the Colorado Reading Project. *Reading and Writing: An Interdisciplinary Journal* 3: 235—248.

Olson, R. K. & Wise, B. (1987). Computer speech in reading instruction, pp. 156—177, in: D. Reinking (ed.), *Computers and reading: Issues for theory and practice.* New York: Teachers College Press.

Olson, R. K. & Wise, B. (1990). Reading whole text with computer-speech feedback for disabled readers. Paper presented at the American Education Research Association Meetings, 16—20 April 1990, in a symposium titled 'Using computer-generated speech to teach reading skills: Current research and future directions'.

Olson, R. K., Wise, B., Conners, F. & Rack, J. (1990). Organization, heritability, and remediation of component word recognition and language skills in disabled readers, pp. 261—322, in: In T. H. Carr and B. A. Levy (eds.), *Reading and its development: Component skills approaches.* New York: Academic Press.

Olson, R. K., Wise, B., 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.

Perfetti, C. A. (1985). *Reading Ability.* New York: Oxford Univ. Press.

Rack, J. P., Snowling, M. J. & Olson, R. I. (in press). The nonword reading deficit in developmental dyslexia: a review, *Reading Research Quarterly.*

Reitsma, P. (1983). Word-specific knowledge in beginning reading, *Journal of Research in Reading* 6: 41—56.

Rosner, J. (1971). *Phonic analysis training and beginning reading skills.* Pittsburgh: University of Pittsburgh Learning Research and Development Center (Publication Series).

Rozin, P. & Gleitman, L. R. (1977). The structure and acquisition of reading, II: The reading process and the acquisition of the alphabetic principle, pp. 55—141, in: A. S. Reber & D. L. Scarborough (eds.), *Toward a psychology of reading.* Hillsdale, NJ: Erlbaum.

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming, *Psychological Review* 96: 523—568.

Spaai, G. W. G., Reitsma, P. & Ellermann, H. H. (1991). Effects of segmented and whole-word sound feedback on learning to read single words, *Journal of Educational Research* 84: 204—213.

Spoehr, K. T. & Smith, E. E. (1973). The role of syllables in perceptual processes. *Cognitive Psychology* 5: 71—89.

Stanovich, K. E. (1986). Cognitive processes and the reading problems of learning disabled children: Evaluating the assumption of specificity, in: J. K. Torgesen & B. Y. L. Wong (eds.), *Psychological and educational perspectives on learning disabilities.* Orlando, FL: Academic Press.

Taft, M. (1979). Lexical access via an orthographic code: The Basic Orthographic Syllabic Structure (BOSS), *Journal of Verbal Learning and Verbal Behavior* 18: 21—39.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games, *Cognition* 15: 49—74.

Wagner, R. & Torgesen, J. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills, *Psychological Bulletin* 101: 192—212.

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

Wise, B. W. (1987). Word segmentation in computerized reading instruction, *Dissertation Abstracts* (University of Colorado at Boulder).

Wise, B. W. (in press). Whole words versus phonics for short-term learning: Comparisons on a talking computer system. *Journal of Experimental Child Psychology.*

Wise, B. W. & Olson, R. K. (1991). Remediating reading disabilities, pp. 631—658, in: J. E. Obrzut and G. W. Hynd (eds.), *Neuropsychological foundations of learning disabilities.* New York: Academic Press.

Wise, B. W. & Olson, R. K. (1992). How poor readers and spellers use interactive speech in a computerized spelling program, *Reading and Writing: An Interdisciplinary Journal* 4: 145—163 (this issue).

Wise, B. W., Olson, R. K. & Treiman, R. (1990). Sub-syllabic units as aids in beginning readers' word learning: Onset-rime vs. post-vowel segmentation, *Journal of Experimental Child Psychology* 49: 1—19.

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.