SESSION VI

INSTRUCTIONAL AND CLINICAL APPLICATIONS

Chris Stout, *President*
Forest Hospital and Foundation

Implementing a long-term computerized remedial reading program with synthetic speech feedback: Hardware, software, and real-world issues

BARBARA WISE, RICHARD OLSON, MIKE ANSTETT, LAURALYN ANDREWS,
MAUREEN TERJAK, VIVIAN SCHNEIDER, JULIE KOSTUCH, and LAURA KRIHO
University of Colorado, Boulder, Colorado

This paper discusses hardware choices, software developments, implementation issues, and preliminary results from an ongoing long-term remedial reading study. Reading-disabled children read books on microcomputers linked to speech synthesizers, obtaining speech feedback on difficult words at whole-word, syllable, or subsyllable levels of segmentation. Word-recognition ability and attitude about reading improved for children using the system. In addition, segmented feedback especially benefited phonological word-decoding skills for most of the children.

Microcomputers equipped with high-quality synthetic speech can provide powerful research and remedial tools in the field of reading disabilities. In our current long-term study, reading-disabled children are improving their deficient word-decoding skills while reading books on microcomputers linked to high-quality speech synthesizers. This paper briefly covers the rationale for the study and then provides an overview of the program as it is run in the schools. Next follow discussions of hardware choices, software developments, and management procedures designed to implement the study and to keep third- to sixth-grade children reading independently for a semester. The paper concludes with preliminary findings from the first semester of the study.

BACKGROUND AND RATIONALE

Reading-disabled, or dyslexic, children find the task of learning to read extraordinarily difficult, despite their nor-

mal intelligence, vision, hearing, and educational background. Population estimates of reading disabilities vary widely depending on the severity of the deficit measured (Olson, 1985). We and many other current researchers have agreed to study disabled readers who meet the above exclusionary criteria (with IQs of at least 90), yet are at or below the 10th percentile in reading.

Most disabled readers' problems relate to language deficits, and not to popularly supposed visual-perception deficits (for reviews, see Liberman, 1982; Vellutino, 1979; Wagner & Torgeson, 1987). Most reading-disabled children show a specific deficit in decoding new and nonsense words, below their general word-recognition ability (Olson, Wise, Conners, Rack, & Fulker, in press; Snowling, 1980). The decoding of new words requires the ability to associate letter groups with speech sounds smaller than whole words, a skill called phonological coding. It constitutes the heritable part of disabled readers' deficient word-recognition skills and correlates strongly with segmental language skills in nonreading tasks (Olson, Wise, Conners, & Rack, in press).

To say that reading disabilities can be inherited is not to say that they cannot be remediated. Some remediated adult dyslexics have learned to read many words by sight, revealing their deficit only when encountering new words in print or when attempting tasks that require manipulating language below the syllable level. Thus, dyslexics can certainly improve their reading ability through extensive

This research was supported by NIH Grants HD11681, HD22223, and HD-2507RR07013-22 to R. Olson. We thank the principals, staff, and students of University Hill, Ryan, and Whittier Elementary Schools for their assistance, and Debbie Aguiar, Rae Sullivan, Nikki Jones, Audrey Torres, and Kathy Mahoney for their help in preparing the materials. We also thank Fran Conners and John Rack for helpful comments and suggestions about analyses. Correspondence may be sent to Barbara Wise or Richard Olson, Department of Psychology, Campus Box 345, University of Colorado, Boulder, CO 80309.

reading practice. It is also quite possible that an inherited deficiency in the ability to learn subword print-to-sound associations can still be improved with powerful remedial techniques. These contentions motivated our study.

In our study, we sought to determine whether reading on microcomputers with immediate speech feedback for word-decoding problems would prove to be a powerful remedial technique for disabled readers. We also wanted to determine whether some form of feedback would particularly improve the phonological deficit that is evident in nonword-reading ability. The students in the study were assigned to different feedback conditions, which determined computer presentation of requested feedback on difficult words. For some children feedback was by whole-word units, for some it was by syllable, and for some, by subsyllable "onset-rime" units (e.g., pl/ant, d/ish). The definitions of the syllable and subsyllable segments were based on earlier studies in which different types of subword units were compared as aids for children's word learning (Wise, 1987; Wise, Olson, & Treiman, 1989). Our goal was to train 250 disabled readers over four semesters. The first phase of the study to be reported here will indicate the general benefits of the system and suggest an overall advantage for one type of feedback. Ultimately, with enough subjects, the long-term study should also enable us to determine whether children with different cognitive and linguistic profiles learn better with one type of feedback or another.

GENERAL METHODOLOGY

General Procedures

The children begin by taking pretests. Pretest measures are used in later calculations of posttest improvement. They also provide information about cognitive and linguistic abilities. Finally, on the basis of word-recognition pretest scores, children are assigned to a directory of stories at an instructional level where they can read fairly fluently but need computer help on 1%–5% of the words.

Each child logs onto the system daily. He or she can choose to hear a brief description of any story title in the assigned directory. The student selects a story and reads for 25 min each day until the end of the semester, using the mouse to "target" difficult words and request help or confirmation from the computer. In three of the four feedback conditions, the computer first highlights and then highlights and pronounces the requested word. The segments are highlighted for an interval on the basis of the following algorithm: highlighted time = (time × number of letters in segment) + 300 msec, where time = 300 msec for whole-word and 400 msec for segmented conditions. The student blends the segments into a whole word, using sentence context to check each attempt. In the fourth, combination, feedback condition, the timing parameters are the same and the words appear on the screen for the same duration as in the other feedback conditions. However, immediate speech and segmentation feedback occur first at subsyllable and then at syllable

levels. Thus, a word in combination feedback has twice the speech feedback, but it does not appear first with speech-free segmentation; this routine was included to encourage students to try to blend segments independently.

To encourage continuous reading for meaning, main idea-comprehension questions are presented at reasonable breaks in the story every 4–9 pages. The student either reads the question and alternative answers independently or can have the computer read them. The student chooses an alternative answer with the mouse, and the program scores the child's response and also highlights and speaks the correct answer, thus providing comprehension feedback. The computer then tests word recognition of five previously targeted words. If fewer than five words have been targeted, the program fills in up to five default words chosen from the text. The words appear one at a time on the screen, highlighted in segments according to the assigned condition, but with no speech feedback. The student reads the word, and then hears the correct pronunciation and self-scores the item.

At the end of the semester, the children take posttests on the reading and language measures. Control students are pre- and posttested, but spend the intervening time on their normal remedial reading and language-arts activities and do not read on the computer.

Equipment and Hardware

The system described above runs on an IBM XT-compatible microcomputer with 640K main memory, a 40-MB hard disk, one 360K floppy drive, a graphics video adapter with monitor, a mouse, a stand-alone speech synthesizer (DECtalk, made by Digital Equipment Corporation), and two serial ports that interface with the mouse and the DECtalk speech synthesizer. The graphics video adapter enables the size and style of the print to be altered to increase legibility. The letters appear twice normal ASCII size, in letter-quality print, with 12 lines of text to a screen and a 13th line for commands.

Programs and materials have been prepared on an IBM AT-compatible computer. All programs have been written in the C language, because of its relative ease of programming and its growing popularity and support. All system and support software has been designed in a modular manner to make the system as hardware-independent as possible. Most of the upper level code would be unaffected if the system were ported to a different host computer. If other components, such as the speech synthesizer or the targeting device, were changed, only the lowest level routines would be affected.

Our original pilot work utilized both Apple computers and IBM PCs, and both light pens and mice (Olson, Foltz, & Wise, 1986; Olson & Wise, 1987). We decided to continue development with XT-compatible computers for several reasons. The XT provides a good environment for software development and has a widely used and supported architecture. It also supports large hard-disk subsystems at reasonable prices. The resulting large amount of storage space allows for flexibility and a large expan-

sion capability. We chose the mouse over the light pen because the light pens had tired some children, they required more repair than the mice, and even our youngest children quickly became proficient with the mouse.

To keep the system as user-friendly as possible, the reading program is menu driven and the mouse is the only input device the children need. Data is stored on the hard disk. The keyboard is used by testers to score tests and to edit and back up files. When the tester is present, the children can use the keyboard to make the synthesizer talk, as an extra motivator. Otherwise, the children interact with the computer only via the mouse. To avoid embarrassment for the readers and disruption for their schoolmates, the children hear all output through earphones. Each system is housed in a table with a lockable shelf large enough to hold all the hardware, leaving only the mouse and monitor on its surface. To minimize distractions when systems are placed in busy rooms, an optional hood screens some tables.

Evaluation of Speech Synthesizer

Our system requires very high-quality speech output. Currently, speech can be produced by a computer in two ways: human speech can be recorded, digitized, and played back through a digital-to-analog (D/A) converter, or artificial speech can be synthesized from text by the computer (Klatt, 1987; Witten, 1982). Synthetic speech requires far less storage space than digitized speech and is flexible enough to allow children's own names to be used in instructions.

From the outset of our project, we have used the DECTalk speech synthesizer, whose "Perfect-Paul"-voice speech we found to be almost as intelligible as tape-recorded human speech (Olson et al., 1986). DECTalk can be interfaced with any computer and will convert ASCII text to pronunciations of English words. DECTalk contains its own internal dictionary with pronunciations of irregular frequent words, often "reduced" to be suitable for presentation within connected text. It also contains phonological rules to generate likely pronunciations of words not in its internal dictionary. Finally, it responds precisely to a phonemic code enclosed in brackets, allowing for the production of almost any desired pronunciation of a word or segment.

Because DECTalk is a high-quality, expensive piece of hardware (currently \$1,700 to educational institutions), we are continuing to search for alternative modes of synthetic speech. We have recently tested the intelligibility of ECHO-II (by Street Electronics) and Smoothtalker (by First Byte), which are less expensive software-based synthetic speech programs that have proven quite adequate in whole-sentence presentations of instructions in some popular educational software. Our system, however, requires single-word and even subword-unit intelligibility.

To evaluate the intelligibility of the software-based systems for single words, 5 adults listened to ECHO-II and 5 listened to Smoothtalker. They heard the same randomly ordered 96 words from the eighth-grade Spache oral read-

ing paragraphs that earlier subjects had heard in a study that compared DECTalk and human speech (Olson et al., 1986). In the earlier study, 8 adults had recognized 96.1% of tape-recorded human pronunciations and 8 had recognized 95.8% of DECTalk's pronunciations of single words. In the present study, subjects recognized only 46% of ECHO-II's and 48% of Smoothtalker's isolated word productions. These percentages translate to error rates of 54% and 52%, which do not differ significantly from each other, but are more than 10 times higher than the 4% error rates of DECTalk and recorded human speech (see Figure 1). Because of these extremely high error rates for adults, we did not test ECHO-II and Smoothtalker further with children as we had done with DECTalk. First Byte is developing new versions of Smoothtalker, which may prove more adequate for isolated word feedback.

Preparation of Reading Materials

To keep third- to sixth-grade boys and girls interested in reading for an entire semester, our program requires a wide variety of reading materials. Students' reading levels range from low first-grade to fifth-grade levels at the start of the program, and to sixth-grade level by the end of the semester. The stories selected cannot depend on illustrations for comprehensibility, since the system does not yet have graphics capabilities. Eighty-one books and 64 stories selected from children's literature on the basis of recommendations from children's librarians, teachers, parents, and students have been prepared so far.

Various measures are used to determine the grade-level difficulty of a book. Once a book has been typed into a file, a program computes its average number of syllables per word and number of words per sentence. Using these two measures, books are assigned to 1 of 10 different directories of stories from low-first- to sixth-grade levels. Suitable ranges for the grade levels were determined by calculating the above measures for the Spache grade-leveled paragraphs (Spache, 1963) and by checking other readability formulas and grade-leveled material (Harrison, 1980). Since disabled readers' deficits lie mainly in

Figure 1. Single-word intelligibility of synthetic and human speech.

word decoding, measures of word difficulty are more important than sentence length for our program. A book's directory assignment is thus determined primarily by its number of syllables per word, and then is adjusted up or down one level for unusually short or long sentences (see Table 1). Directory assignments are also lowered for extremely repetitive language or raised for complex story structure or vocabulary. Some assignments have now been further adjusted due to unusually low or high rates of targeting or comprehension scores in Phase 1 of the project. New books continue to be added as children or teachers suggest them.

All words in all books used need to have segment divisions marked and clear pronunciations of word and subword units coded. We have developed much software to facilitate this process. The most interesting of these programs enables us to code and store a word only once in a dictionary. Once a word has been coded, a story's textfile need only contain each word's dictionary address and position on the screen. If a child uses the mouse to request help on a word, the word's segmentation and pronunciation information can be retrieved in one single access of the dictionary. Memory storage needs for the stories themselves are minimal compared to our earlier work, when textfiles contained segmentation, screen placement, and pronunciation information for every word.

The dictionary itself is an indexed file that can be searched quickly for the correction or listing of entries. Because the index file is larger than available random access memory, the index structure and search algorithms have been optimized for performance in a disk-based memory system, using a binary tree structure of order 10. The current dictionary contains more than 17,000 word entries in only five levels.

To prepare the materials, books are typed in 40-character lines and desired page breaks are indicated with pluses. Coders determine logical breaks in the story, and write a main idea question at each break. Coders write three alternative answers and indicate the correct answer in the raw file so that the program can score children's responses appropriately. At lower reading levels, questions occur, on average, every four screens; at higher levels, they occur, on average, every seven screens.

Table 1
Measures Used to Assign Books to Directories

Directory (Level)	Syllables/Word (Range)	Words/Sentence	
		Avg.	SD
Low First	1.00-1.39	6.69	.684
High First	1.40-1.49	7.09	1.15
Low Second	1.50-1.59	7.34	1.09
High Second	1.60-1.64	8.05	1.17
Low Third	1.65-1.69	8.61	2.08
High Third	1.70-1.75	9.61	2.09
Fourth	1.75-1.79	9.63	1.64
Fifth	1.80-1.89	11.36	2.69
Sixth	1.90-1.97	12.40	4.80

A program compares newly typed books with the dictionary, placing unknown words in a file. Trained coders edit these word files, marking segment divisions and coding pronunciations in temporary files. "Dictionary experts" run a second-stage editing program to check and correct the work and enter the word into the dictionary. This allows coders to have hands-on practice while learning our rather complex segmentation rules and our own modifications of DECtalk's phonemic codes, while ensuring that the feedback for every word will be coded as clearly and consistently as possible. Words that have different pronunciations and meanings depending on context, such as "bow" or "record," are marked as ambiguous. This creates a second entry that is coded for the alternative segmentation and pronunciation information.

After the new words have been entered in the dictionary, a formatting program checks margins, inserts needed screen breaks, and converts the input into a textfile containing information about number of pages, number of lines per page, and each word's line position and dictionary address. Finally, a disambiguating program is run on textfiles. The story is written to the screen as it will appear to the reader, but with ambiguous words highlighted. A coder chooses the appropriate alternative for the ambiguous word, and the program writes that alternative's address to the textfile.

READING WITH ORTHOGRAPHIC AND SPEECH SEGMENTATION (ROSS): PHASE 1

We began the first phase of our long-term project with third- to sixth-grade disabled readers from three elementary schools in Boulder, Colorado. Teachers and administrators selected students who they thought were of average intelligence but who read at or below the 10th percentile for readers in their school. School personnel scheduled the students to read 25 min at least four times a week, mainly during times that would otherwise be spent on remedial reading, classroom reading, or language arts.

We were not ready to begin pretesting until late February, and we had lower attendance than we had anticipated, due to Spring break, annual achievement testing, end-of-year special events, and confusions in some schools where personnel had scheduled students at three or four different times during the week. Therefore, the students spent an average of only 10 h on the system (besides the 3 h of pre- and posttesting), which was less than we had expected.

Method

Subjects

Sixty-two third- to sixth-grade children were selected from among children recommended by Whittier and University Hill Elementary Schools in Boulder and by Ryan Elementary School in Lafayette. Recommended children whose verbal or performance IQ estimates were below 90 were dropped from further consideration for par-

ticipation in the project. [These estimates were based on our administration of the Vocabulary, Similarities, Picture Arrangement, and Block Design subtests of the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974), using a conversion formula recommended by Sattler (1982).] A few other children did not qualify because their tested reading ability turned out to be well above the 10th percentile. Three subjects dropped out of the study, and one moved. The remaining sample comprised 21 girls and 37 boys.

Materials

The reading materials included the 81 books and 64 stories coded for segmentation and speech feedback, with questions and word-recognition tests as described in the General Methodology section above.

Procedures

Pretests. Before they began to read, the subjects took various tests over 5 days, as described in the General Methodology section. The tests were interspersed with attitude questionnaires, with games, and with training on the system. Descriptions follow of the word-recognition and nonword-decoding tests whose results are discussed in this paper.

Two measures of word recognition were included. The Peabody Individualized Achievement Test for word recognition (Dunn & Markwardt, 1970) was adapted for the computer. One word at a time appeared on the screen. The subject read the word, the tester scored the final response as correct or incorrect, and the next word then appeared on the screen. The test stopped when the subject missed 5 of the last 7 items.

The children also took a timed word-recognition test that was based on one used in our twin study (Olson, Wise, Conners, & Rack, in press). Each word appeared on the screen for 2 sec, the student read the word aloud, and the tester scored the final response. The subjects read a screener list until they missed 4 items; then they were placed in a 220-item word list on the basis of that score, and read on until they missed 10 of the last 20 items.

A non-word-reading test was given on 2 different days, with 43 items each day. The nonwords ranged from one-syllable items containing two phonemes and three letters to two-syllable items with eight phonemes and nine letters. The items appeared on the screen and remained until the tester scored the student's final response as correct or incorrect by either grapheme-phoneme correspondence rules or by rhyme analogy to real words (e.g., "tive" could be pronounced to rhyme with either "give" or "hive").

Training, reading, and monitoring. The subjects were assigned to feedback conditions: whole-word, syllable, subsyllable, or a combination condition of first subsyllable and then syllable feedback. The conditions were balanced for age and phonological-decoding ability as indicated by the nonword pretests.

At the beginning of training, the subjects were assigned to a level one directory below their PIAT word-recognition grade-equivalent score, so that the reading level would be easy while they were first learning to use the system.

During training, testers monitored the children's oral reading. If a student completed a sentence without targeting a misread word, the tester pointed to that word to remind the child to target it. The tester wrote down independently the targeted and reminded words so that the percentage of words reminded on monitored days could be calculated. The directory assignments were repeatedly adjusted so that the child would target at least one word per screen; thus, word-recognition checks generally contained targeted, and not default, words. Because many disabled readers have become so used to guessing that they are unaware of when they have misread a word, the students were encouraged to target both unknown words and any words that had caused a struggle.

After 2 weeks, the students read independently, using the system as described earlier in the General Methodology section. The tester returned weekly to monitor progress, adjust directory assignments, and give stickers for books read. The data files on monitored days were flagged, so that the targeting rates and percent correct on comprehension questions and word-recognition tests could be compared for monitored and independent days.

Posttests. At the end of the semester, the earlier tests were administered again as posttests to both experimental and control subjects, and the attitude questionnaire was given again to students who had read with the system. A written questionnaire was also delivered to their teachers so that we could get their feedback and improve our school procedures.

Results

When time for hearing instructions, selecting stories, answering comprehension questions, and doing word-recognition checks is subtracted from the daily reading times, the 10-h average reading time for all experimental subjects shrinks to just over 5 h. Twelve children read stories for less than 4 h each, and so were dropped from the analyses to be reported here. Analyses of the entire group show the same pattern of results for posttest improvement. Significant effects remain, but are weaker. However, comparisons of monitored and independent days are meaningless for the 12 short-time readers, because they had so few independent days. The 36 students used in the results reported below averaged 6.7 h of actually reading stories, or about 13 h on the system (plus the 3 h of pre- and posttesting).

Daily Reading Behavior

Analyses of monitored versus independent days were done by *t* tests. One system did not flag the monitored sessions, but the data for the other 28 students whose monitored sessions were flagged can be examined. The means for these data are shown in Table 2.

Table 2
Monitored versus Independent Reading Behavior

Behavior	Monitored	Independent	<i>p</i>
Number of targets in 25 min	33	27	> .1
Percent correct word recognition	78	87	< .001
Percent correct comprehension	95	89	< .01

Table 3
Targeting Rates per 25 Minutes Across Feedback Conditions

Feedback Mode	Rate	<i>SD</i>
Whole-word	26.5	19.1
Syllable	27.0	18.4
Subsyllable	19.3	8.3
Combination	45.7	26.6

The students self-scored their daily word-recognition tests 9% higher on independent days than they did on days when they were monitored by testers [$t(27) = 4.24$, $p < .001$]. Also, the students' scores on the computer-scored comprehension checks were 6% lower on independent than on monitored days [$t(27) = 3.58$, $p < .01$]. Analyses of variance were conducted with feedback condition as the independent variable and percent correct on word-recognition checks and comprehension questions as dependent variables; there were no effects of feedback condition in either analysis.

The students targeted six more words per 25-min session on monitored than on independent days, but this difference was not significant [$t(26) = 1.65$, $p > .1$]. Analysis of variance showed that students in the combination feedback condition had significantly higher rates of targeting than did students in the other conditions, although the difference had not been significant in the unselected sample (see Table 3). (All other differences discussed in the results section that were significant in the selected sample had also been significant in the unselected sample.) This greater difference may have been due to the methodological difference described earlier. Combination feedback had immediate speech with no initial speech-free highlighting to encourage independent blending attempts; the children may have liked the immediate speech but may have paid less attention to the segments in this condition than in the other feedback conditions. Children in the combination feedback mode also had nearly significantly lower initial word-recognition scores (PIAT, $p = .055$; timed recognition, $p = .07$), due partly to chance and partly to the loss of students who were dropped from the sample. Lower level children take more time to read words, read fewer words per session, and have higher targeting rates. The targeting rates of the other segmentation conditions did not differ significantly from each other.

The children's behavior did not change significantly over the semester with respect to targeting rate or daily checks of word recognition and comprehension. However, the students advanced an average of 3.2 story directories during that time, reflecting improving reading ability.

Feedback conditions did not differ significantly on directory advancement, although a trend favored subsyllable feedback by 1 directory [4 directories advanced for subsyllable vs. 2.9 and 3 for the others; $F(3,32) = 1.77$, $p < .2$]. The percentage of targeted words that required a reminder from the tester before the child targeted them on monitored days also did not change significantly over the semester, averaging 19.8% and not differing by feedback condition [$F(3,18) = .73$].

Posttest Improvement in Word Recognition

Analyses of variance were performed with posttest minus pretest difference scores as the dependent variable and feedback mode as the independent variable. Analyses of variance were also conducted with pretest and posttest scores as repeated test measures. The levels of significance for all significant effects were the same in the repeated-measure and the difference-score analyses. Difference score analyses will be reported, since their means are easy to interpret. Results for timed word recognition, untimed word recognition, and nonword decoding can be seen in Figures 2, 3, and 4.

Feedback condition showed a significant effect in improvement for timed word recognition [$F(4,40) = 5.1$, $p < .01$]. Planned contrasts were calculated using Dunn's test (Bonferroni t ; Howell, 1982). Experimental subjects gained more than did controls [$t'(41) = 2.55$, $p < .05$]. Students with subsyllabic feedback did not gain significantly more on this test than did students with whole-word feedback [$t'(41) = 1.42$], but students reading with whole-word feedback did achieve more gains on this measure than did the control subjects [$t'(41) = 2.34$, $p < .05$], as expected. However, on this measure, combination feedback did not yield more gains than the control condition [$t(41) = .95$].

There was no significant effect of feedback condition on untimed word recognition (PIAT) [$F(4,41) = 1.09$,

Figure 2. Time-limited word recognition: Number of words gained.

Figure 3. Untimed word recognition (PIAT): Gain in grade equivalence.

Figure 4. Nonword reading: Gain in percent correct.

$p = .38$]. The average gain on the PIAT test was only 3.8 words, with means ranging from 2.6 words for the controls to 5 words for the children in the subsyllable condition. This measure may not contain enough words to be sensitive to differential gains over short periods of time. When the words gained on both word-recognition tests are summed for one analysis, the result shows that the gains in word recognition are more than two times greater for the experimental subjects than for the controls [$F(4,40) = 5.24, p < .01$].

Feedback condition showed a significant effect on improvement in non-word-decoding ability [$F(4,38) = 3.6, p < .02$]. Planned contrasts revealed that experimental subjects improved more than control subjects [$t'(38) = 2.99, p < .05$], children in segmented conditions improved more than students in whole-word feedback [$t'(38) = 2.81, p < .05$], and subsyllabic feedback yielded more gains than whole-word feedback [$t'(38) = 2.82, p < .05$]. However, for nonword reading, students reading in the

whole-word condition did not improve more than control subjects [$t'(38) = .65$].

Attitude Changes

The questionnaire data yielded positive results. Both before and after the training period, experimental subjects had answered questionnaires concerning attitude about reading and about most favorite and least favorite subjects in school. Control subjects were not questioned, so control versus experimental analyses were impossible.

A tally of the questions, however, indicated an exceedingly positive response. Of the 30 children who were given the open-ended questions about most and least favorite school subjects, 9 showed some positive change in attitude about reading. Others didn't mention reading either before or after training. Twenty-one of 35 students said they were reading more at the end of the program than they had before the program started. With direct questioning, 33 children said they liked to read more at the end of the study than they had before the study began; only 2 said that their feelings about reading were the same before and after the study.

Responses from the students' 10 classroom and resource teachers were also extremely positive. Nine of the 10 said they had seen general improvements in most of their students' reading abilities, attitudes about reading, and attitudes about school. The 10th teacher had only 1 student on the system who had very low attendance due partially to a very confusing schedule. This student had made no gains either in reading ability or attitude, and his teacher rightly judged that the system had not helped that student and that his scheduling was disruptive. The teachers said they did not have time for closer involvement with the system themselves; they indicated that having a tester come weekly was very motivating and helpful for the students.

Discussion

In general, the results do support the contention that microcomputers equipped with synthetic speech can provide a helpful remedial tool for reading-disabled children. Analyses of daily reading behavior indicated that the children were able to use the systems nearly as well on their independent days as on days when a tester helped monitor their reading and targeting behavior. Targeting rates did not differ significantly regardless of whether or not a session was monitored. Scores on self-scored word-recognition tests and computer-scored comprehension checks did differ, but only by less than 10%. Targeting rates and daily comprehension and recognition scores stayed fairly stable over the training period, with general improvement reflected in all students' advancing to more difficult directories of books.

Experimental subjects showed significant improvements in comparison with control subjects in timed word recognition, despite relatively short time on the system. Segmented and whole-word feedback did not differ from each other in terms of gains on timed word recognition. Differ-

ences on the standardized untimed PIAT word-recognition test were in expected directions, but were not at all significant, reflecting wide variability within conditions due to the small number of words gained. Each word gained on the PIAT counts for about .2 reading-grade levels, so the measure appears to be insensitive to differential gains over less than one semester.

Interpretation of the gains made by experimental subjects over controls on the posttest word-recognition test is clarified by knowing how many of the words gained were actually targeted for feedback at some point in the study. Because students picked books of interest at different grade levels, this overlap varied widely for all subjects. In analyzing this overlap, "words gained" means those which had been misread on the pretests but successfully read on either posttest word-recognition test. Only 12% of the words gained had actually been targeted during the study, and this amount of targeting overlap did not differ according to condition (whole-word, 15%; syllable, 7%; subsyllable, 14%; combination, 13%; $p = .74$). The other 88% of the words gained had thus *not* been previously targeted. This could mean that they had been seen and learned through reading with context in the books on the computer or during regular classroom instruction. The more interesting possibility is that the feedback conditions helped subjects improve phonological-coding skills that could support generalized independent word recognition on the posttests, and indeed this possibility is supported by the data from nonword reading.

Data concerning improvement in the phonological-coding skills reflected in reading nonwords reveals that experimental subjects improved significantly more in this skill than did the control subjects. Even more interestingly, students in segmented feedback conditions achieved significantly larger gains on this measure than did students who had received whole-word feedback. Segmentation had not shown a significant advantage over whole-word feedback on word recognition. These results suggest that reading with the support of computer speech feedback in any feedback mode improves word recognition, but that segmented feedback especially improves the phonological coding skills evident in the significant gain in nonsense-word reading. The improvement in generalizable phonological coding skills helps students to read novel words more effectively, even without ongoing computer assistance.

The combination feedback condition differed from the other conditions in two ways, and the children improved significantly less with it, at least on timed word recognition, despite significantly higher targeting rates. Children in this condition may have been less active in decoding attempts than were children who first saw segments highlighted without feedback. The children usually figured out the word on the basis of the first spoken feedback and were already reading the rest of the sentence when the second highlighting and speaking of syllable feedback occurred. Thus, they may have actually spent only about half as much time concentrating on the unknown word as the other children had.

Taken together, the results from the first phase of the project suggest that the system can indeed provide a powerful remedial reading tool. Teachers and students seemed very positive, despite scheduling and attendance problems. We have made adjustments in our procedures to simplify scheduling and to streamline the time spent on and between other modules of the program, in order to maximize the time students spend in actually reading stories. We hope that results from more subjects with more hours will replicate the significant benefits of the computerized reading system, and will clarify questions about differences among the feedback conditions for children with differing cognitive and linguistic profiles.

REFERENCES

- DUNN, L. M., & MARKWARDT, F. C. (1970). *Examiner's manual: Peabody Individual Achievement Test*. Circle Pines, MN: American Guidance Service.
- HARRISON, C. (1980). *Readability in the classroom*. Cambridge, England: Cambridge University Press.
- HOWELL, D. C. (1982). *Statistical methods for psychology*. Boston: Duxbury Press.
- KLATT, D. (1987). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America*, **82**, 737-793.
- LIBERMAN, I. Y. (1982). A language-oriented view of reading and its disabilities. In H. R. Myklebust (Ed.), *Progress in learning disabilities* (Vol. 5). New York: Grune & Stratton.
- OLSON, R. K. (1985). Disabled reading processes and cognitive profiles. In D. Gray & J. Kavanagh (Eds.), *Biobehavioral measures of dyslexia*. Parkton, MD: York Press.
- OLSON, R. K., FOLTZ, G., & WISE, B. (1986). Reading instruction and remediation with the aid of computer speech. *Behavior Research Methods, Instruments, & Computers*, **18**, 93-99.
- OLSON, R. K., & WISE, B. (1987). Computer speech in reading instruction. In D. Reinking (Ed.), *Computers and reading: Issues for theory and practice*. New York: Teachers College Press, Columbia University.
- OLSON, R. K., WISE, B., CONNERS, F., & RACK, J. (in press). Organization, heritability, and remediation of component word recognition and language skills in disabled readers. In T. Carr & B. Levy (Eds.), *Reading and its development: Component skills approaches*. New York: Academic Press.
- OLSON, R. K., WISE, B., CONNERS, F., RACK, J., & FULKER, D. (in press). Specific deficits in component reading and language skills: Genetic and environmental influences. *Journal of Learning Disabilities*.
- SATTTLER, J. M. (1982). *Assessment of children's intelligence and spatial abilities*. Boston: Allyn & Bacon.
- SNOWLING, M. J. (1980). The development of grapheme-phoneme correspondence in normal and dyslexic children. *Journal of Experimental Child Psychology*, **29**, 294-305.
- SPACHE, G. (1963). *Diagnostic reading scales*. Monterey, CA: McGraw-Hill.
- VELLUTINO, F. (1979). *Dyslexia: Theory and research*. Cambridge, MA: MIT Press.
- WAGNER, R. K., & TORGESON, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, **101**, 192-212.
- WECHSLER, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corp.
- WISE, B. W. (1987). *Word segmentation in computerized reading instruction*. Doctoral dissertation, University of Colorado at Boulder.
- WISE, B. W., OLSON, R. K., & TREIMAN, R. (1989). *Subsyllabic units as aids in beginning readers' word learning: Onset-rime versus post-vowel segmentation*. Manuscript submitted for publication.
- WITTEN, I. (1982). *Principles of computer speech*. New York: Academic Press.