Journal of Speech and Hearing Disorders, Volume 55, 635–655, November 1990

TABLETOP VERSUS MICROCOMPUTER-ASSISTED SPEECH MANAGEMENT: RESPONSE EVOCATION PHASE

LAWRENCE D. SHRIBERG JOAN KWIATKOWSKI TEREZA SNYDER
University of Wisconsin—Madison

This is the third in a series of studies on the use of microcomputers with speech-delayed children. Two repeated-measures designs (n = 15) and five case studies were completed to compare tabletop management at early and late stages of the response development phase with two comparable, computer-assisted drill-and-practice activities. Discrimination of correct articulatory responses was mediated by the clinician in all modes, rather than by speech recognition hardware, but all contingent reinforcement in the computer modes was presented by animation graphics. The two computer modes were identical except for the addition of fantasy involvement in one of the modes. Findings indicated that the three modes of intervention were equally effective, efficient, and engaging. Subject-level analyses suggested that microcomputer software has excellent potential to engage children in drill-and-practice for late-phase response evocation, when the target sound is stimulable, but limited usefulness with young children at early-phase response evocation, when specific articulatory behaviors need to be cued. Discussion considers learning, child, and hardware/software factors in microcomputer-assisted speech management.

KEY WORDS: articulation, phonology, microcomputers, intervention, response development.

The present report is the third in a study series (Shriberg, Kwiatkowski, & Snyder, 1986, 1989b) exploring learning factors, child factors, and hardware/software factors in computer-assisted management for children with developmental phonological disorders. The following review introduces individual considerations in each of these three domains, followed by an integration of concepts leading to the research questions.

Learning Factors in Computer-Assisted Speech Management

Borrowing from theoretical and applied literatures in education, psychology, and in particular, motor skills learning, intervention programs in communicative disorders typically are divided into discrete phases of learning. Figure 1 is a representation of four of the most typical phase-level frameworks for management, as reflected in texts and synthesis papers on speech management (e.g., Bernthal & Bankson, 1988; Costello, 1984; Creaghead, Newman, & Secord, 1989; Hoffman, Schuckers, & Daniloff, 1989; Shelton & McReynolds, 1979). As shown in the bottom row, the simplest diachronic view of alternative intervention models is a one-phase view in which no distinctions are made among phases of learning. Two-phase learning models typically posit different mechanisms for the acquisition of new behaviors and the transfer of behaviors to relevant linguistic and social contexts. Three-phase models add intervention activities aimed at

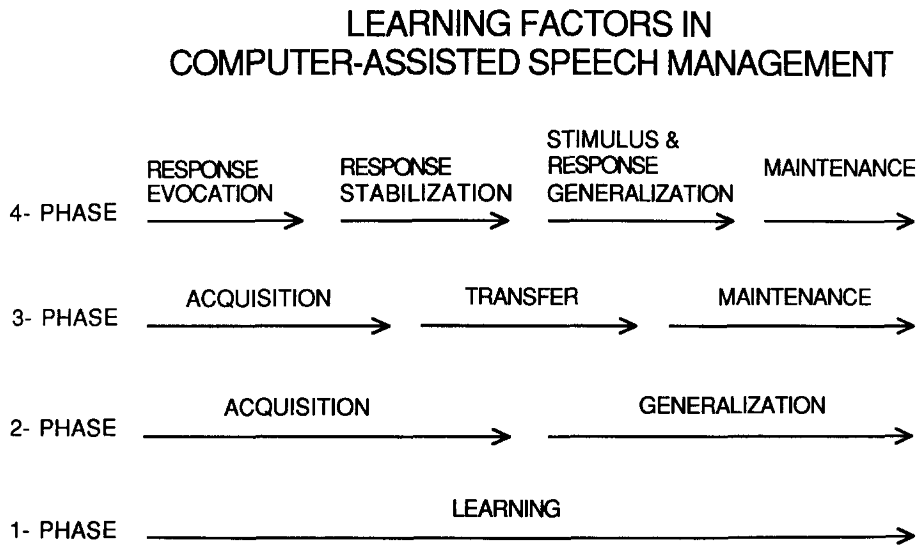

FIGURE 1. Learning factors.

  0022-4677/90/5504-0635$01.00/0

the maintenance of generalized behaviors. Finally, a four-phase model divides acquisition and transfer into subordinate phases, including response evocation, response stabilization, response and stimulus generalization, and maintenance.

Compared to the number of studies of generalization in speech management, only a small literature is centered specifically on the response development phase of management. Most of the clinically useful literature remains in the form of classic and updated "approaches" and lists of "tricks" to get children to say sounds correctly. Reduced interest in research study of this phase of learning could plausibly be due to at least two evident problems: the basic difficulty of accomplishing controlled research designs on such targets and trends away from the study of residual articulatory errors. Although most contemporary intervention procedures in child phonology do invoke two-, three-, and four-phase models of learning, the tasks of evoking new articulatory behaviors generally is left to the speech-language pathologist's clinical expertise.

Child Factors in Computer-Assisted Speech Management

A second domain requiring consideration in computer-assisted management is those individual difference variables associated with the children for whom programs are designed. Figure 2 includes five variables that must be accounted for in computer-assisted speech management. The two speech variables, phonological pattern and metalinguistic awareness, reflect current focus on both linguistic analyses of output phonology and the cognitive-linguistic organization underlying the error pattern. Research on error pattern typologies and studies of metalinguistic variables are oriented toward efficient target selection, posing alternatives to the traditional criteria for target selection and sequencing based on the developmental order of speech sound mastery. It is important to underscore the inherent difficulty in teaching children both the articulatory features underlying correct sound production and the organizational aspects that govern such phonological variables as allophonic and morphophonemic variation. Compared to other targets of computer-assisted learning, such as the recognition of orthographic symbols, math skills, and certain social and language concepts, teaching the correct articulation of speech sounds by computer is a formidable challenge.

Individual differences in the three other causal correlates listed in Figure 2—motivation, etiological background, and cognitive-learning style—are also sources of variance in all forms of teaching, including computer-assisted management. Children who are not inherently motivated to improve speech obviously are good candidates for anything that might improve their attention to each of the elements in the clinical process. Due to the failure, to date, of studies seeking etiological subgroups among children with developmental speech delays, the impact on management of such etiological factors as auditory-based causal backgrounds compared to subtle speech-motor based backgrounds remains essentially unstudied. Finally, as suggested in Figure 2, individual differences in children's cognitive styles must be accounted for in the materials and pace of learning. Perhaps the largest single obstacle to the development of computer-assisted materials for speech management is to provide for branching routines that accommodate the diverse cognitive-learning styles of young preschool-age children who currently qualify for speech management services. Each of these five child factors will be addressed in more detail in later discussions.

Hardware and Software Factors in Computer-Assisted Speech Management

Figure 3 is an attempt to identify the significant hardware and software variables for computer-assisted speech

CHILD FACTORS IN
COMPUTER-ASSISTED SPEECH MANAGEMENT

SPEECH

PHONOLOGIC PATTERN
METALINGUISTIC AWARENESS

CAUSAL CORRELATES

MOTIVATION
ETIOLOGIC BACKGROUND
COGNITIVE-LEARNING STYLE

FIGURE 2. Child factors.

HARDWARE AND SOFTWARE FACTORS IN COMPUTER-ASSISTED SPEECH MANAGEMENT

		SOFTWARE		
		ANALOGIC	ICONIC	THEMATIC
	KEYBOARD/TOUCH			
HARDWARE	SIGNAL PROCESSING			
	LINGUISTIC RECOGNITION			

FIGURE 3. Hardware and software factors.

management. The nine cells in this schema should be useful for classifying the distinctive features of each speech-management program in an eventual microcomputer-assisted technology.

Considering first the *hardware* variables, the input to computer-assisted speech management can be by keyboard or some other transducer, by signal recognition technology, or by speech recognition technology. Most current special educational software uses standard ASCII keyboard input, rather than signal processing peripherals, although the use of touch screens, switches, and other input devices is common. The second alternative, some type of signal processing input, includes devices that are sensitive to acoustic (e.g., VisiPitch™) and tactile (e.g., Hardcastle, Jones, Knight, Trudgeon, & Calder, 1989) stimuli. What is only emerging at present are inexpensive hardware and software packages that perform speech recognition of various levels of complexity (e.g., SpeechViewer™; Watson, Reed, Kewley-Port, & Maki, 1989), typically requiring talker-specific training to develop the relevant stimulus-response contingencies.

Among the many ways to classify *software*, the schema in Figure 3 is based on the cognitive-affective processing the software evokes from the user. The most difficult cognitive task, one that generates low levels of affective engagement, occurs when the program displays an acoustic analogue of the input signal. Hence, as shown in studies that provide lissajous figures, wave forms, and other analogues of vowel and consonant signals, such information can be difficult for young or cognitively involved children and may not be inherently motivating. Displays that translate signals into icons attempt to meet both needs, using such methods as indicating changes in loudness or pitch by the relative size or color of some friendly object, such as a clown's nose or bow tie (e.g., SpeechViewer™). Finally, some software organizes stimuli in thematic ways, as in games, simulations, and other engaging activities that relate individual trials to one another. The potential power of such software with young children is evident in the popularity of microprocessor-controlled video games, in which children's cognitive and affective resources evidently can be engaged and sustained in isolation from other human interactions for exceedingly long periods.

Response Development and Computer-Assisted Speech Management

From the foregoing review of associated issues, it is clear that response development may be the most difficult phase of speech management at which to attempt computer-assisted intervention. First, because children may or may not readily be stimulable for a target sound, it is necessary to divide response evocation (see Figure 1) into two subphases. Early-phase response evocation activities are necessary for children who need extensive models and articulatory cues from the clinician to produce close approximations of the target sound. Late-phase response evocation activities are sufficient for children who need only an imitative auditory model or no model to produce a correct approximation of the target sound. Whereas the stabilization phase of response development is oriented toward response consistency in complex and meaningful, but structured, linguistic environments, these two subphases of response evocation involve repeated trials of simple, nonmeaningful isolated sounds, words, and phrases. Sustained activities of this sort, particularly for children at early-phase response evocation, require that the child is motivated to remain attentive. Whether the management agent is a clinician, a computer, or a clinician aided by a computer, and whether or not the computer program has speech-recognition input, response evocation requires three elements: reliable discrimination of the relevant auditory and gestural behaviors for all phoneme targets, the ability to branch to alternative and appropriate subroutines, and sensitivity to the motivational-affective processes that subserve attention and metaphonologic engagement.

These perspectives, together with findings from two prior studies comparing speech-delayed children's responses in tabletop modes to computer-assisted versions of the same activities, can be interpreted to predict limited efficacy for computer-assisted response evocation. In one study series (Shriberg et al., 1986), speech-delayed children's responses to booklet-based articulation testing were compared to computer-assisted articulation testing of comparable stimuli. Results indicated that although children's accuracy of articulation did not differ statistically by presentation mode, computer-assisted testing was associated with better attention and task persistence. The second series (Shriberg et al., 1989b), which studied microcomputers at the stabilization phase of response development, also did not yield quantitative support favoring the effectiveness of either intervention mode. However, there was support for the conclusion that the computer mode was more effective for some children due to its higher engagement value, as attested on several variables in three studies within the series. In computer mode, children tended to be more frequently and appropriately looking at the computer than at the clinician or elsewhere. Moreover, 83% of the children stated that they would prefer to work in computer mode rather than tabletop mode in a future management session.

The potential problems for computer-assisted response evocation suggested by these prior results concern the relevant attentional focus at each of the two subphases described above. Even if speech recognition technology were available for judgments of correct sounds, the relevant feedback cues for shaping correct target sounds need to be provided by clinicians attuned to children's articulatory and motivational behaviors. Hence, particularly at the earliest phase of response evocation, when children's correct sound production requires a complex of antecedent cues, this phase of learning would seem to be difficult to program for computer-assisted management. Specifically, computer-assisted response evocation activities might be counterproductive because the engagement values of the computer would vie for attention with the face-to-face activities needed to model and evoke phonetic behaviors. For the later phase of response evocation, in which trials are conducted for children who are able to produce correct sounds with only an auditory cue, computer-assisted management might be found more effective because the children could hear the models while looking at the computer. Moreover, the added engagement might engender and maintain attention and focus on the target behavior.

An integrated solution to each of the variables considered to this point is to consider the possible use of fantasy in computer-assisted response development. In a discussion of motivational factors in computer-assisted learning, Lepper (1985) has emphasized the significant role of *fantasy,* as well as *challenge, mastery,* and *perceived control.* Conceivably, the power of video games to compel sustained attention and skills acquisition could be achieved in speech-management software that also invoked fantasy involvement in thematic ways. Specifically, software could be built to include fantasy elements that would encourage a child to consult the clinician for all response evocation cues.

The current study series includes two controlled studies and one controlled clinical validity study to test two questions at the response evocation stage of response development: (a) Are there statistically significant differences in the *effectiveness, efficiency,* or *engagement* value of computer-assisted response evocation activities compared to tabletop mode? (b) Is the inclusion of fantasy involvement in response evocation associated with gains in effectiveness?

METHOD

Subjects

Demographic data and speech characteristics for the 20 children in Studies 1, 2, and 3 are summarized in Table 1. Five additional children, 3 in Study 1 and 2 in Study 2, were excluded because technical problems affected one of the experimental modes, or the target was stabilized prior to administering all modes. All children were attending a university clinic for developmental phonological disorders. Each of the three studies was conducted during a different academic semester.

As shown in Table 1, the 2 girls and 7 boys in Study 1 ranged in age from 2:11 (years:months) to 6:5 (M = 4.4). The age range for the 1 girl and 5 boys in Study 2 was 4:2–7:5 (M = 5:3). In Study 3, the 2 girls and 3 boys ranged in age from 3:7 to 8:2 (M = 5:5). All children were from middle-class socioeconomic backgrounds and were judged within normal limits for speech-hearing mechanism function, cognitive-linguistic level, and psychosocial development, based on screening in the local public schools prior to referral to the university clinic. Speech status, based on percentage of consonants correct (PCC) (Shriberg & Kwiatkowski, 1982; Shriberg, Kwiatkowski, Best, Hengst, & Terselic-Weber, 1986), ranged from mild-moderate to severe. All children were classified as speech delayed, based on their error patterns as described by computer-assisted natural phonological process analysis (Shriberg, 1986). Substitutions were the primary error type for most of the children, with 2 children having primarily deletion errors and 4 children having primarily distortion errors.

Speech Targets

At least 1 of the 20 children was assigned one of the following speech targets: k, f, s, ʃ, tʃ, l, ɝ and the clusters

TABLE 1. Demographic data and speech characteristics for the 20 children in Studies 1, 2, and 3.

Study	*Subject*	*Gender*	*Age (years:months)*	*Speech status*[a]	*Primary error type* Deletions	Substitutions	Distortions
1	1	F	2:11	MM		X	
	2	M	3:4	MM		X	
	3	M	3:10	MS		X	
	4	M	4:1	MM		X	
	5	F	4:3	MM		X	
	6	M	4:4	MM		X	
	7	M	5:1	MM		X	
	8	M	5:4	MM			X
	9	M	6:5	MM			X
2	1	F	4:2	S	X		
	2	M	4:2	MM		X	
	3	M	4:3	MM		X	
	4	M	4:4	MS	X		
	5	M	6:7	MS		X	
	6	M	7:5	MM		X	
3	1	M	3:7	MM		X	
	2	M	3:11	MM		X	
	3	F	5:10	MM		X	
	4	M	6:3	MM			X
	5	F	8:2	MM			X

[a]MM = mild-moderate; MS = moderate-severe, S = severe.

sp, st, and k. To avoid potential ceiling effects across and within experimental training sessions, the single target selected for each child met each of the following three criteria. First, the sound was the most difficult for the child to produce from among his or her current evocation targets, as determined by current degree of success in therapy. Second, the sound was never produced correctly during two spontaneous speech probes. One probe, a continuous speech sample that included a minimum of five word types containing the target, was administered prior to the first experimental training session. A second probe, a carrier phrase-level task, was administered prior to each experimental session. Five words containing the target sound were produced within a carrier phrase (e.g., "Put (word) on the table") in the context of a game (e.g., the child directed the clinician to place the word cards in a designated place in anticipation of a hiding-finding game).

The third criterion met for each target sound was linguistic context level, which was determined immediately before administering the training task in each experimental training mode. To determine the appropriate level, a 30-item imitative probe was used. It consisted of 10 isolated *sound* trials, 10 different *syllable* items, 5 *word* items, and 5 different words embedded in a *carrier phrase*. The highest linguistic level at which the child obtained at least 20% correct, but not more than 40% correct with an imitative auditory model, was selected as the level for training. All decisions were made on-line by the child's clinician and confirmed by the clinician's supervisor.

STUDIES 1 AND 2

Tasks and Materials

Table 2 is a description of the agents for the instructional and motivational events in the three training modes

TABLE 2. Agents for the instructional and motivational events in the three training modes in Studies 1 and 2.

Training mode	*Agent*	*Instructional events* Task instructions	Auditory models and cues	Verbal knowledge of results	*Motivational events* Visual knowledge of results	Primary reinforcement
Tabletop	Clinician	X	X	X	X	X
	Computer	–	–	–	–	–
Computer	Clinician	X	X	X	–	–
	Computer	–	–	–	X	X
Computer + robot	Clinician	–	X	–	–	–
	Computer	X	–	X	X	X

els tested in Studies 1 and 2. Each training mode consisted of repeated cycles of practice on the target followed by a reinforcement period. As indicated in Table 2, the clinician provided the auditory models and additional cues as needed to evoke production of the target in each of the three training modes. In tabletop mode, the clinician also provided instructions and visual and verbal knowledge of results, and the primary reinforcer, reading sections of a story. One of two popular children's stories (Zion, 1958, 1965) was assigned in a randomized counterbalanced order to each child. The clinician's involvement in providing instructions, knowledge of results, and reinforcement was different in each of the computer modes. In computer mode, the clinician provided instructions and verbal knowledge of results; whereas in the computer + robot mode, a friendly robot figure named "PEP" provided these through synthesized speech. In both computer modes, visual knowledge of results was presented via graphics animation, and reinforcement was delivered by PEP the robot, who narrated a story about his adventures. A different story was assigned on a random counterbalanced order to each computer mode.

Figure 4 includes representative materials from the practice periods in the three training modes. Each panel is a black-and-white screened photograph of the original materials for the tabletop mode and computer + robot mode, respectively. In tabletop mode, plastic chips were used to indicate the number of trials per trial block. A paper frame signaled the current trial. Apart from the movement of the frame for the next trial, there were no visual displays to signal response accuracy. The graphic displays in the computer and the computer + robot modes were identical, except that the computer + robot graphics (shown in Figure 4) included the robot, PEP. In both computer modes, the number of trials in the trial block was represented by magicians' hats. A flashing frame signaled the current trial. When a child said a target sound correctly, a rabbit or colorful bouquet of flowers appeared out of the hat. Sounds that were close approximations of the target yielded only a partial display of the rabbit or flowers (in Study 1 only). Sounds said incorrectly were signaled by a short beep and yielded no display. Articulatory judgments were made unobtrusively by the clinician, who discreetly depressed the appropriate key on the computer keyboard.

Computer programming was accomplished in 6502 machine language by an artist-computer programmer using a 128K Apple IIe microcomputer and the Echo IIb™ speech synthesis system. The software was developed so that findings would be generalizable to the lower end microcomputer technology that is available currently in the schools where most speech-delayed children are served. The computer modes were presented on a 128K Apple IIe microcomputer equipped with a 65C02 microprocessor, a new monitor ROMS, and an 80-column AppleColor Composite Monitor IIe.

Procedures

The three experimental training modes were administered in a randomized counterbalanced order during three successive 50-min management sessions. All training sessions were conducted by the child's clinician, who was familiar with the child's current speech production levels and learning style. Eleven master's level student clinicians (6 in Study 1, 5 in Study 2) were trained to follow the scripted protocol presented in Appendix A to assure similar and consistent task administration in all training modes. The protocol contained procedures that were routine in the university clinic during evocation training (cf. Shriberg, Kwiatkowski, & Snyder, 1989a).

Four changes from the Study 1 protocol were made in the protocol for Study 2. They are depicted in bold type in parentheses in Appendix A. All changes were motivated by the clinicians' and the children's reactions to the protocol used in Study 1. One change was to reduce the

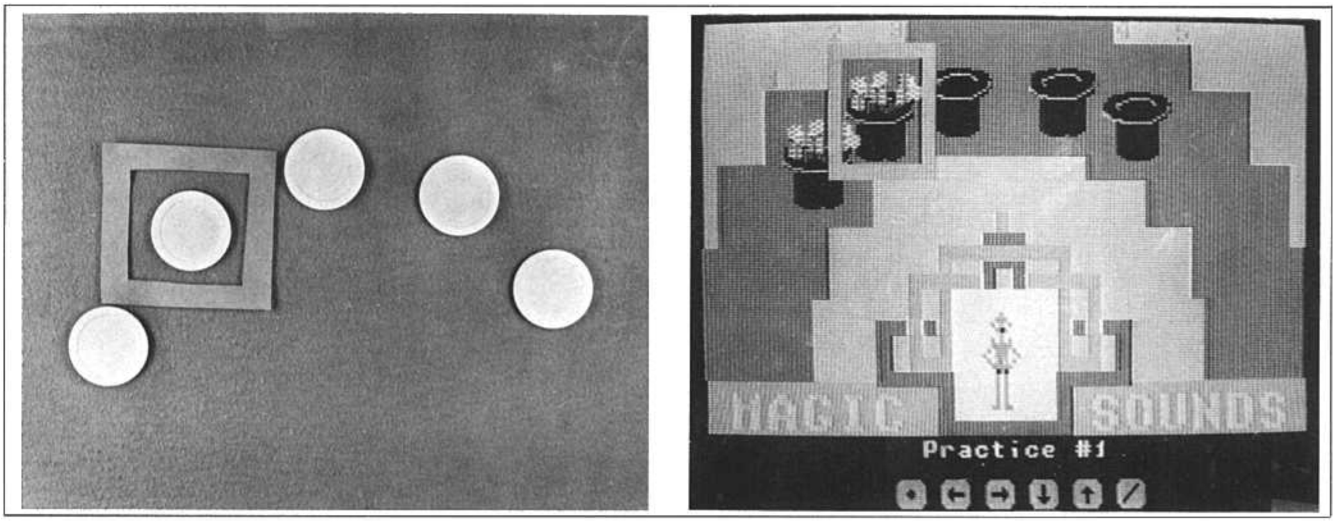

FIGURE 4. Representative materials from the two intervention modes. The left panel displays the practice materials used in the tabletop mode. The right panel displays comparable materials for the computer + robot mode. A graphic display similar to the display in the right panel, without the robot in the practice screen, was used in the computer mode.

number of trials per trial block from 10 to 5 because both clinicians and children reacted negatively to the length of each trial block. To maintain the same total number of trials (50), the number of trial blocks was increased from 5 to 10. A second change was to replace the partial display for "close" in the computer modes with a full display because the partial display in Study 1 was judged too punishing for some children. Thus, in Study 2 both "correct" and "close" approximations of the target received the same visual feedback display. The third change in Study 2, because the children seemed unnaturally passive throughout the training sessions in Study 1, was to have the children physically active during the reinforcement period by letting them depress the appropriate computer key or turn a page to see part of a story. The fourth change was made because the phrases used to provide knowledge of results for incorrect responses in Study 1 may not have suggested to the children that they had the option to actively seek the clinician's help to correct errors. To suggest the option of more active involvement in learning the target, the original evaluative phrases for incorrect responses were changed to the prompt "ask for help" in the computer + robot mode and a parallel comment "let me help you" in the computer and tabletop modes in Study 2.

In both Study 1 and Study 2, children were allowed up to three attempts per trial to achieve an acceptable response. Acceptable responses included both correct and close approximations of the target as individually defined for each child. For all children, the first attempt to produce the target was always in response to an auditory model. Cues for the second and third attempts were individualized, but were not adjusted on a per-response basis trial to trial. To prevent the child from becoming frustrated by repeated unsuccessful attempts to produce an acceptable response, response definitions for the third attempt virtually guaranteed an acceptable response.

All experimental training modes and probes were administered in the same therapy room with the computer always present. Children sat to the left of the clinician and on the same side of the table. Although most of the children had had some exposure to the computer in the clinic, none had experience with the tabletop and computer versions of the tasks under study. For the period of the study, children practiced the target only while in a training mode.

Administration of all training modes was simultaneously audio- and videotaped. Speech probes were only audiotaped. All audio recordings were obtained using a Marantz PMD201 audiocassette recorder with matching Sony EC-3 microphone and TDK audiocassettes. Video recordings were obtained on ¾-in. 3M UCA videocassette tapes using a Panasonic WV-6000 color camera feeding a Sony VO600 videocassette recorder housed in an adjacent room. The operator of the camera, who was hidden behind a screened partition in the therapy room, consistently maintained upper body pictures of both child and clinician while filming the training interaction. A second computer monitor, located directly behind the child and clinician, allowed simultaneous filming of materials being presented on the computer screen. Clinicians maintained the children's lip-to-microphone distance at approximately 15 cm. Because audiotape and videotape recording was a standard part of the management routine, the children were indifferent to the presence of the recording equipment.

After each training session, the clinicians annotated their qualitative impressions of the children's behaviors and the clinician-child interaction. Following the last management session in each study, to assess their perception of the training modes, the children were asked which, if any, of the three modes they would prefer for a future session.

Reliability

Interjudge reliability for judgments of children's articulatory accuracy was obtained between the clinicians' on-line judgments and the second author's judgments from audiotape. A randomly selected 10% sample of each child's data (excepting 1 child in Study 1 for whom accurate judgment required visual information and 1 child in Study 2 who spoke too softly to allow valid coding of responses from the audiotape) yielded average point-to-point agreement for correct, close, and incorrect judgments of 97%, with a range of 81%–100%.

Data Reduction

The children's behaviors were coded from videotapes for three intervention constructs that were also coded in an earlier study (Shriberg et al., 1989b). These constructs, termed *Effectiveness*, *Efficiency*, and *Engagement*, were operationally defined for the current study to reflect the behavioral domains described in the coding protocol in Appendix B.

Effectiveness represents the frequency with which attempts to produce the target were judged correct, close, or incorrect by the clinician. Hence the judge did not determine response accuracy but rather coded articulatory responses according to the clinician's evaluations on the videotape.

Efficiency reflects the length of time spent in practice-reinforcement units during the experimental training session. Units included repetitive cycles of preresponse (for presentation of instructional models and cues), response, and postresponse (for presentation of knowledge of results) within each practice period, and a reinforcement period following each practice. The judge used specific verbal and behavioral cues from the administering clinician to identify each practice-reinforcement unit. There were 5 units per training session in Study 1 and 10 in Study 2.

Engagement depicts the affective aspects of the therapy situation, reflected in the interaction of the child and the clinician with each other and with task materials. The three engagement variables, as described in Appendix B, were eye gaze, facial expression, and verbal responses.

The clinician's eye gaze served as the reference for judging the appropriateness of the direction of the child's gaze. Behaviors that could not be judged were coded "cannot judge."

Instrumentation, Training Procedures, and Coding

The data reduction instrumentation used in Studies 1 and 2 was similar to that described in Shriberg et al. (1989b). All videotapes were played on a Sony V02610 videotape deck and Panasonic CT201M 20-in. color monitor with all start-pause functions controlled by custom software running on a Commodore Vic 20 microcomputer (VCR CONTROLLER, Epp, 1987). A computer-aided behavioral analysis system (TERMITE, Ver Hoeve, 1986) permitted on-line coding directly into a Harris/800 minicomputer. TERMITE classified and stored each occurrence of a behavioral category (i.e., the Effectiveness and Engagement codes) and timed all categories classified as "continuous" (i.e., the Efficiency codes) to the nearest 100 ms. The three behavioral variables subsumed under Engagement were coded on a 12-s time sampling schedule. After each 12-s Viewing period, a status light mounted above the video monitor signaled a 1-s Judging period. The videotape then automatically stopped, still-framing the screen. After an 8-s Coding period, the tape restarted to begin another sampling cycle. Verbal behaviors were judged as they occurred throughout the 1-s Judging period. All other engagement behaviors were judged on the basis of the still-framed screen image at the end of the Judging period.

The three judges in the current study were familiar with the instrumentation and coding procedures because they had been the judges for a similar study at the stabilization phase of management (Shriberg et al., 1989b). Each judge was a second-year master's level student in communicative disorders and was familiar with clinical procedures and process. Prior to coding tapes for the current study, the judges had trained for 3–5 hr per construct coding each of the three intervention constructs. Additional training was conducted to clarify and refine changes made in the original codes to increase coding sensitivity. Changes included (a) the addition of separate codes for reporting response accuracy for up to three attempts per trial for Effectiveness, (b) the elimination of redundant posture behaviors from Engagement coding because Posture could not be coded independent of Gaze and Facial Expression, and (c) the addition of codes to existing Facial Expression and Verbal codes for Engagement coding. Training was accomplished for one construct at a time using pilot tapes of 4 different children.

After training and the interjudge reliability assessment described below were completed, each judge coded all 45 videotapes (Study 1: 9 children × 3 taped conditions; Study 2: 6 children × 3 taped conditions) for a single construct—the same construct each had judged in the stabilization-phase studies. All coding for Study 1 was completed within a 1-month period. Two months later, coding of videotapes for Study 2 was begun and completed as the tapes were obtained over a 3-month period.

Validity and Reliability of Coding

To assess the concurrent validity of Engagement coding, the judge who coded the Engagement construct completed a questionnaire after coding each videotape to (a) annotate her clinical impressions of the child's behaviors and the child-clinician interaction and (b) identify whether the assigned codes accurately reflected her clinical impressions. The judge's responses on the questionnaire indicated a 100% correspondence between global codes and clinical impressions. For 41 of the 45 sessions (91%), the questionnaire data indicated that the engagement codes captured the judge's specific impressions of the child's level of engagement and the clinician-child interaction. Similar concurrent validity tests were not necessary for Effectiveness and Efficiency, each of which had inherent face validity. That is, percentage of correct trials is the standard measure of training effectiveness, and accumulated time is a standard measure of efficiency.

To assess coding reliability across subjects and modes, each of the experimental training modes for 3 randomly selected subjects from Study 1 was assigned in a randomized counterbalanced order to one of the intervention constructs of Effectiveness, Efficiency, and Engagement (3 children × 3 modes = 9% of the data in Study 1). Interjudge reliability studies for each construct were initiated 1 week after the completion of training for each intervention construct prior to coding the data. The intrajudge reliability was accomplished by comparing codes assigned during the original data reduction to codes assigned during the interjudge reliability assessment. Intrajudge estimates of coding stability were available only for Effectiveness and Efficiency. Intrajudge reliability estimates were not obtained for Engagement coding because limitations in the instrumentation did not allow exact freeze-frame replication of each of the original judging segments.

Reliability data for all comparisons were calculated by means of a utility program (RELYONME, Olson, 1987) that produced point-by-point agreement percentages for original-rejudge comparisons. Average interjudge reliabilities were: Effectiveness, 94% (range 92%–97%); Efficiency 92%, (range 90%–94%); and Engagement, 90% (range 89%–92%). Average intrajudge reliabilities were: Effectiveness, 96% (range 95%–99%); and Efficiency, 97% (range 95%–98%). These intra- and interjudge reliability values are consistent with values obtained for the comparable set of codes in the stabilization-phase studies.

Study 3

Study 3 was designed as a set of five case studies to assess whether the controlled experimental protocols used in Study 1 and Study 2 might somehow be limiting the sensitivity of the dependent variables to potentially real group-level effects in Effectiveness. Clinicians' annotated comments in both Study 1 and Study 2 indicated that the experimental training protocol prevented them from responding both technically and interpersonally to the moment-to-moment learning and affective needs of the children. Hence, methods selected for Study 3 were designed to be more consistent with clinical practice, or what has been referred to experimentally as a variant of loose training methods (Stokes & Baer, 1977).

Task and Materials

The task and materials used in Study 3 were essentially similar to those used in Studies 1 and 2 except for the following three changes. First, on the assumption that differences in the visual displays used during practice negatively affected outcomes, a tabletop display was developed that was exactly analogous to the rabbits/flowers display on the computer screen. Second, because different reinforcers might differentially motivate attention and task performance for different children, the primary reinforcers were individualized for each child. Finally, to avoid potential problems related to the intelligibility of the synthesized speech, only tabletop and computer modes were compared.

Procedures

Procedures for Study 3 were similar to procedures used in Study 2 with the following changes. The two experimental training modes were administered by the same experienced clinician (JK). Each mode was administered twice in rotation during the same 50-min management session. Instructions at the beginning of the task were presented verbally, but not demonstrated. In addition, the training task was limited to two trial blocks, and the number of trials per trial block was individualized for each child. At the end of each trial block the child received a star on a scorecard. When both trial blocks were completed children received the primary reinforcer. Most important, in this revised protocol the clinician was given the freedom to adjust instructional cues and linguistic levels on a per-response basis during training, depending on the child's immediate performance and personal needs. In addition, the clinician was free to provide the child with knowledge of performance in addition to knowledge of results when judged necessary. As in Study 1 and Study 2, all task modes were simultaneously audio- and videotaped.

RESULTS AND DISCUSSION

STUDY 1 AND STUDY 2

In consideration of sample size and the nonnormal distributions, the Friedman Analysis of Variance by Ranks was selected for the repeated measures comparisons. An alpha level of .05 was selected for all comparisons, in consideration of goals at this exploratory stage of research. Prior to proceeding with the mode comparisons, nonsignificant Friedman tests indicated the absence of significant main effects for order of mode administration.

Effectiveness

Figure 5 is a display of the Effectiveness data for Studies 1 and 2. Results for mean percentage correct first-try responses summed over trial blocks are presented at the left, with the average correct combined with close responses shown at the right. As suggested by the large and overlapping standard deviation bars, no significant group differences emerged in either study on either dependent variable. Inspection of individual data for each child suggested that approximately half the children had notably higher average percentage correct first-try responses in one mode or another, whereas the remaining children had similar scores in all modes. These individual data are shown in Figure 6.

Efficiency

Figure 7 is a display of the Efficiency data for Studies 1 and 2. Data are presented as average duration in seconds for the three components of the response evocation cycle—the clinician's preresponse instructions (models and cues), the period of the response, and the clinician's postresponse knowledge of results and instructions. With the diversity across children again indicated by the large standard deviation bars, there was only one significant effect, as shown by the asterisk for postresponse in computer + robot mode in Study 2 ($\chi^2 = 9.0$; $df = 2$; $p < .011$). On average, the postresponse period in the computer + robot mode was approximately 3 s longer than in tabletop mode. As in the two earlier studies in this series (Shriberg et al., 1986, 1989b), the trend for the computer modes to take longer was associated primarily with the time needed to present computer graphics and additionally, in the current study, with hardware constraints that prohibited simultaneous presentation of graphic displays and synthesized speech.

Engagement

The four panels in Figure 8 reflect children's engagement in the three modes. The few behavioral events that were coded "cannot judge" are excluded from the average percentage occurrence data presented in each of the panels.

Gaze. The top two panels include the percentage of occurrences for which children's eye gaze was matched appropriately to the clinician or to the materials, or was looking elsewhere. As is evident in the left-most variable, in both Study 1 and Study 2, the child's gaze was more

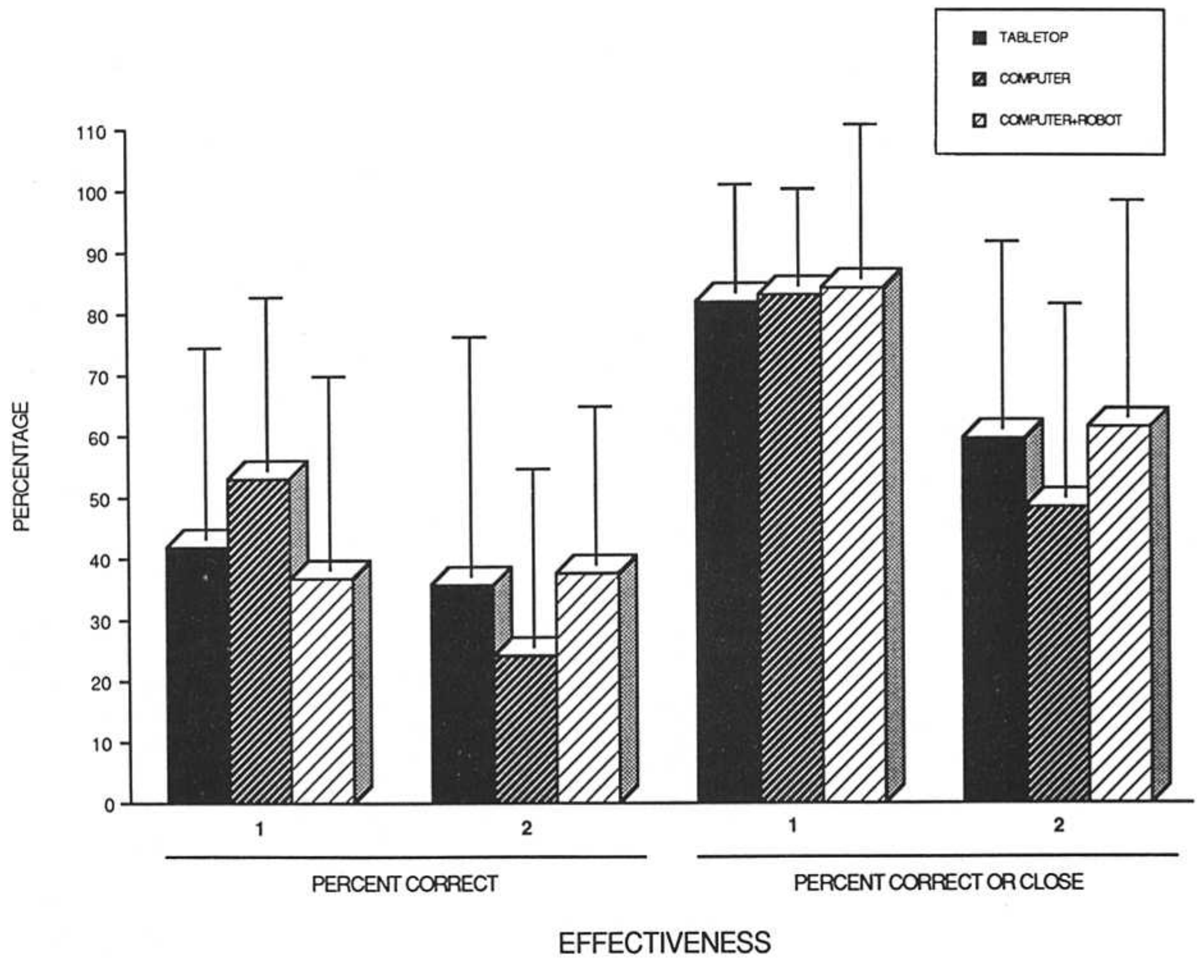

FIGURE 5. Summary of the Effectiveness data for Study 1 and Study 2. The height of the bars indicates the mean performance for each variable; the small bars indicate one standard deviation from each mean.

often matched to the clinician during tabletop mode than it was in the computer modes. The difference was statistically significant in Study 1 [$\chi^2(2) = 14.889$, $p < .001$]. For the middle variables, children's gaze was less often matched to materials in tabletop mode, with differences statistically significant in both studies [Study 1: $\chi^2(2) = 13.556$, $p < .001$; Study 2: $\chi^2(2) = 8.333$, $p < .012$]. Finally, as shown in the right-most set of bars, trends in both studies were for the children more often to be looking elsewhere in the tabletop time samples, with the mean difference approaching statistical significance in Study 2 [$\chi^2(2) = 6.33$, $p < .052$].

Verbal behavior. The lower left panel in Figure 8 is a display of the second set of Engagement codes, here reflecting children's verbal behaviors. As shown, throughout each sample neither the percentage of verbal behaviors coded as positive nor the few behaviors coded as negative differ significantly across modes.

Facial expression. The lower right panel in Figure 8 includes the summary data for facial expression, the third index of Engagement. As shown in the left set of bars, in most sampling points children had neutral facial expression. Although the data in the right set of bars suggest a trend for more frequent occurrences of positive facial expressions in the computer modes, there were no statistically significant differences among the three means. Negative facial expressions occurred only a few times in the hundreds of sampling points.

Child preferences. Results from a fourth index of Engagement are shown in Table 3, which provides subjective data on the three training modes from the children's and the clinicians' perspectives. When asked which mode they would like to try again, 9 of the 12 children (75%) for whom these data were available wanted to continue with one of the computer modes, with computer + robot highly favored (67%). Consistent with children's expressed preferences, as shown in Table 3, clinicians' annotated impressions of the children's behavior during each training mode indicated that 11 of the 15 children (73%) were judged more engaged in the computer modes than in tabletop. The perceived engagement value of the computer modes was nearly evenly divided between computer and computer + robot.

In addition to an expressed preference for computer + robot mode, records indicated that children demonstrated their interest in PEP the robot in several other ways.

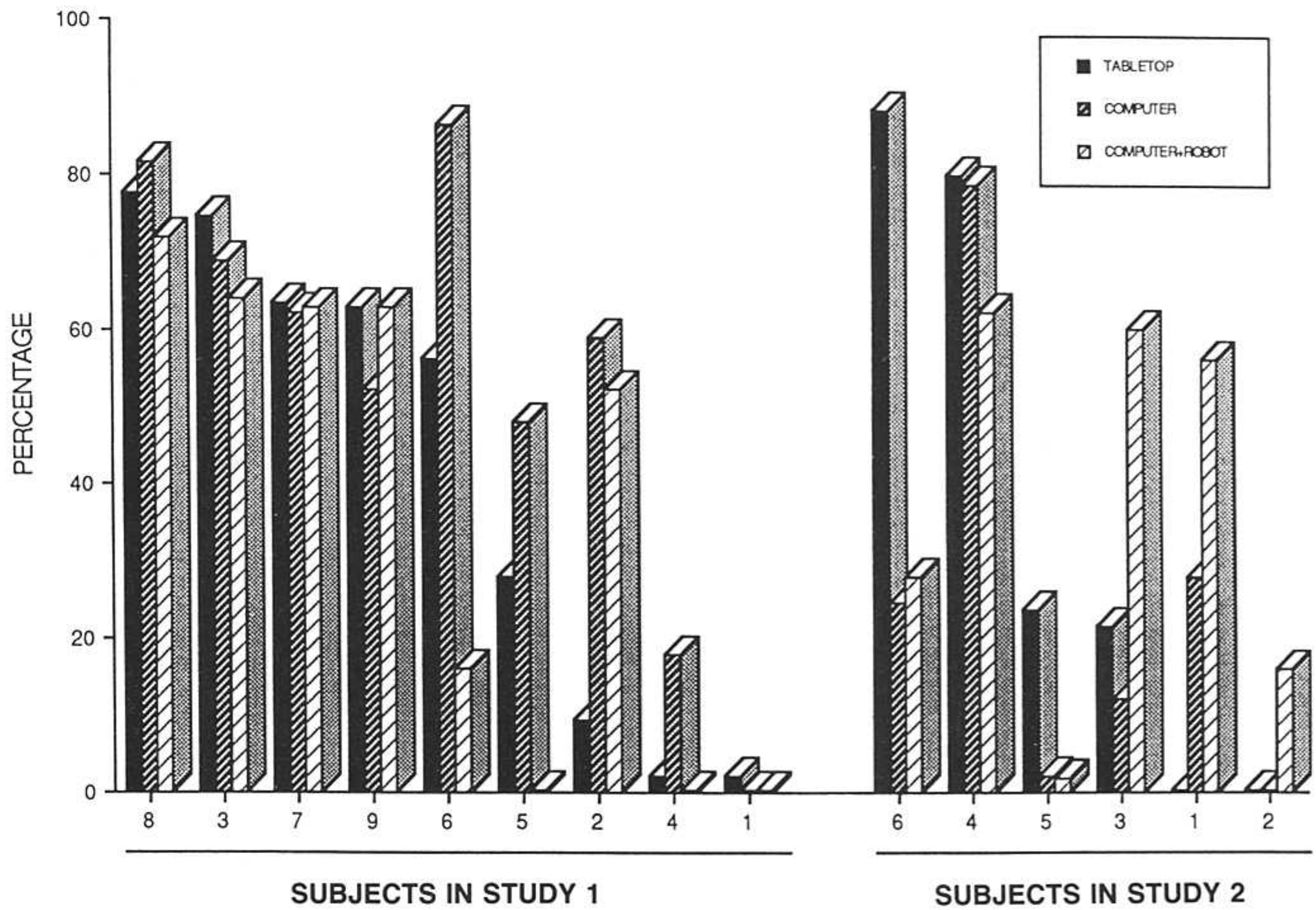

FIGURE 6. Individual subject Effectiveness data in the three training conditions. Subject scores within each study are sorted from left to right in descending order of percentage correct in tabletop mode.

Some of the children asked questions about PEP's experiences, friends, and habitats; others imitated PEP's voice and feedback phrases during self-initiated home practices; others routinely responded to PEP's question regarding their readiness to practice at the beginning of each practice period; and one child asked his mother to stop at the videotape store to see if they could rent a movie about PEP.

Reasons given by some of the children for selecting a mode suggested that the children did not always choose the mode they found most interesting or enjoyable. Some children rejected even personally preferred modes because they recalled the speech task as more difficult, found the partial display punishing, or had difficulty understanding the synthesized speech. Other childen chose preferred modes in spite of negative experiences. For example, of the 4 children who found the partial display punishing, 2 selected a computer mode when given a choice. There was no evidence in the group or individual data to suggest that mode preference was related to how successful the child had been at producing the target when training in the mode.

STUDY 3

Results of the 5 case studies were similar to the results in Study 1 and Study 2, lending additional clinical support to findings in the two experimental studies. Although 4 of the 5 children (80%) indicated a preference for the computer mode, the number of correct or close evocation trials did not differ between the computer and tabletop modes. The 3 children who had the most difficulty in response evocation had variable performance that was independent of training mode. The 2 children who readily could produce the target sound with specific instructional cues made steady improvement over the course of the repeated experimental training modes, with rate of progress also apparently independent of mode. Although all 5 children were visually more engaged by the materials on the computer, they appeared to use the visual displays in both modes only as a pacing device to monitor movement toward completion of the task and claiming their reinforcer. They also appeared to be trying as hard to produce the target during both modes; when the task was too difficult, as it became especially for 1 of the children, neither the computer nor tabletop mode was powerful enough to keep the child involved and working.

DISCUSSION

Consistent with the two prior study series, the present

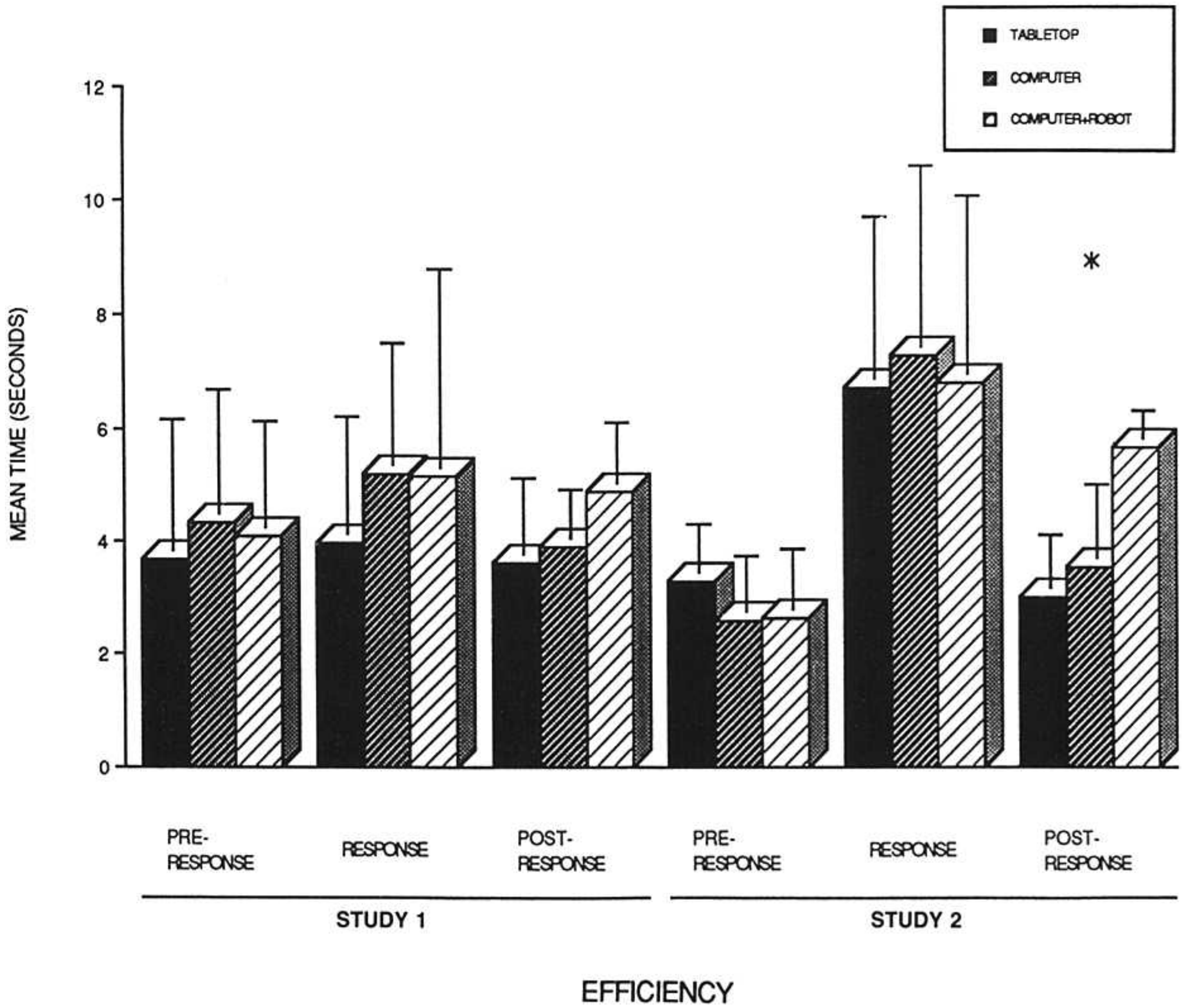

FIGURE 7. Summary of the Efficiency data (means, standard deviations) for Study 1 and Study 2.

findings suggest that neither of the two forms of computer-assisted management was clearly more effective, efficient, or engaging than tabletop mode. Although tabletop mode was associated with more frequently matched child-clinician gaze, perhaps enabling a child to better profit from the clinician's training cues, this mode was also associated with more gazing elsewhere (i.e., not at the clinician or at task materials) and with more off-task verbalization than occurred in at least one of the computer modes. Such data, combined with the child preference findings, suggest that the computer modes may be more engaging for at least some children. As in the stabilization studies, individual differences were the rule, with child-level performance, self-report, and clinical impression data suggesting that some children were engaged substantially by one or another of the modes. Although most children expressed a preference for the computer modes and demonstrated interest and involvement with PEP the robot, this mode was no more effective than the other two modes in evoking correct or close responses. Thus, the results in each of the three study series indicate that computer-assisted modes are not associated with better articulation responses in articulation testing, response stabilization, and, now, response evocation. The following discussion considers several technical and design issues for continued research in microcomputer-assisted speech management.

Hardware and Software Factors

Clearly, a major limitation to an effective computer-assisted management technology involves the level of microcomputers and input/output peripherals available to clinicians in public schools. Findings in these three study series document the influence of technical considerations on the types of software that can be written for computer-assisted speech management on these platforms. Essentially, the clarity and speed of video and audio materials are important factors for program efficacy and efficiency. In the first study series, limitations in the quality of graphics on the Apple IIe were associated with reduced recognition of some of the articulation test stimuli. In the second study series, the reduced speed of presentation of

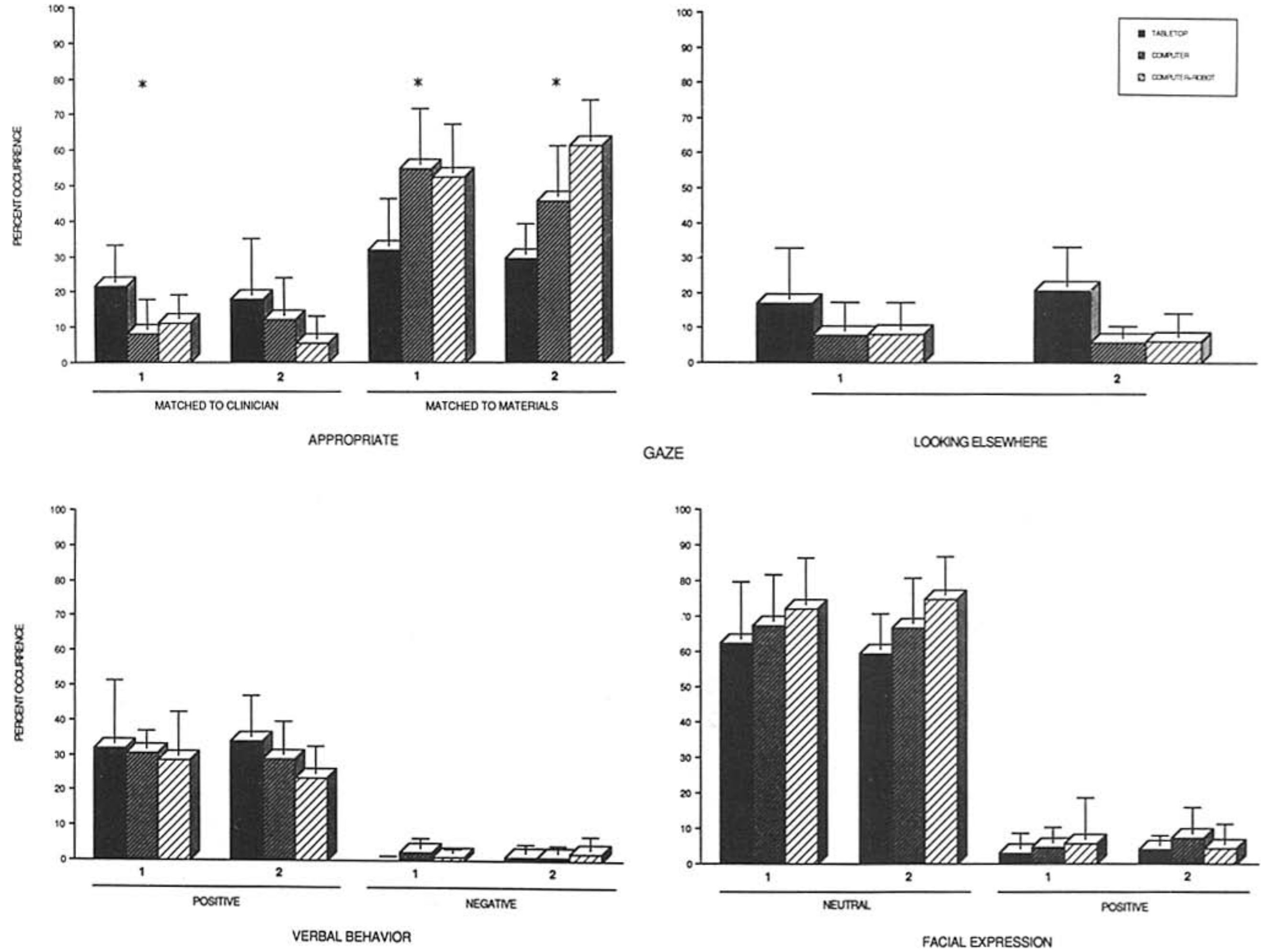

FIGURE 8. Summary of the Engagement data for Study 1 and Study 2, including Gaze (top panels), Verbal Behavior (lower left panel), and Facial Expression (lower right panel).

some graphics elements negatively affected the efficiency of computer-assisted response stabilization. In the present series, the time needed for certain graphics presentations also was associated with a statistically significant efficiency difference, although the average difference of only a few added seconds is essentially of little clinical significance.

The most important hardware issue in the present study was the speech synthesis, which was designed to play a key role in establishing the fantasy figure, PEP the robot. The design attempted to test whether the child's clinician could be used as a resource in response evocation, with the fantasy figure, PEP, telling children to ask for help from their teacher when they needed assistance to correct incorrect responses. In fact, children in Study 2 never asked the clinician for help following wrong responses, even when the robot invited them to do so. Therefore, it could be concluded that this third study series did not really test the efficacy of involving children in a fantasy figure and taking active participation in learning.

In the present context, the focus is on why children did not ask their clinicians for help. A probable reason is that the robot's synthesized speech was too difficult to understand. Although the Echo IIb™ (and previous models) appears to be the most frequently used speech synthesis device in instructional software due to its early availability to software developers and its low cost, studies indicate that its intelligibility is lowest among other devices. High-end systems have intelligibility percentages in the 80% range, several systems score in the mid to low 50% range, and the Echo models have been associated with intelligibility percentages approximating 30% (Keating,

TABLE 3. Children's preferences for each of the three training modes in Study 1 and Study 2. Preference data were available for only 12 of the 15 children.

	Mode preferences			
Source	*None*	*Tabletop*	*Computer*	*Computer + robot*
Child	1 (8%)	2 (17%)	1 (8%)	8 (67%)
Clinician	0 (0%)	4 (27%)	6 (40%)	5 (33%)

Evans, Wyper, & Cunningham, 1986; Kelly & Chial, 1988). Other discussions in microcomputer efficiency studies have stressed the need for instructional phrases to be intelligible, meaningful, and nonrepetitive (Grover, 1986; Heywood, 1986; Massey, 1988). Thus, PEP's instructions at the beginning of the task may not have been understood, and the children may not have realized that they were to take the responsibility to ask for help. PEP's prompt, "ask for help," which occurred throughout the task, may not have been appreciated as an opportunity to ask the clinician for assistance, or perhaps the prompts were not perceived as real communication. Clinicians were instructed to monitor for children's difficulty understanding the synthesized speech and restate the message and then repeat each synthesized speech prompt as necessary, but the children may not have always communicated that they were not understanding. More generally, children may have a high tolerance for low-intelligibility speech when delivered through any type of electronic or audiovideo media, such as microcomputers.

Learning Factors

Several findings appear to document the importance of phase-level learning concepts on the potential efficacy of computer-assisted speech management. In the present study, the subdivision of response evocation into an early and a later phase was important, reflecting respectively a child's need for complex antecedent cues versus only auditory models. At the earliest phase of learning, it is important that the learner be provided with error-specific information (i.e., knowledge of performance, as well as knowledge of results) (Blume-Cohen, 1985; Steinberg, 1984). In Study 3, the inclusion of knowledge of performance, that is, specific information about exaggerated or extraneous tongue postures, appeared to have a facilitative effect on subsequent attempts to produce the target.

Child Factors

As in most training studies in this discipline and others, individual differences in children's cognitive and motivational needs may be the most important source of variance. With only three sessions and no baseline controls, including the more clinically flexible procedures in Study 3, the group designs did not allow for individual functional analyses of the efficacy of the three training modes. Several individual differences among the child factors listed in Figure 2 were evident in the quantitative findings and anecdotal observations.

One unexpected finding was the reaction of some children in Study 1 to the partial screen displays that indicated less than accurate articulatory responses. Similar scoreboards that display a cumulative record of trial-to-trial performance are standard in clinical practice, with the expectation that they provide knowledge of results and motivate sustained effort. The behaviors of some children, however, indicated that they perceived the display of partial success as aversive. As described for subjects in the token loss study reported by McReynolds and Huston (1971), children in the present study showed signs of negative reactivity, avoiding eye contact and engaging in a variety of off-task behaviors. These were more frequent in the computer modes than in tabletop mode, suggesting that potentially aversive computer-associated stimuli may have more salience for children. Specifically, as in the first study series in which children were more compelled to finish the articulation test in computer mode, computer-based stimuli and consequences may be perceived as more powerful than tabletop materials. Although only positive reinforcement was used in these studies, it could be that some children experience some types and schedules of aversive consequences as especially powerful when delivered by a microcomputer. The effect seems consistent with the annoyance adults experience when computer programs signal incorrect user actions by auditory beeps and repetitive visual displays indicating "wrong."

On the positive side, the magnitude of engagement generated by computer software in speech training is also clearly an individual difference variable. For certain children, fascination with the computer graphics appeared to provide a strong motivational component. Thus, in the present study, tabletop and computer modes may each have been effective with certain children for different reasons. The computer mode may have been effective due to its engagement value, which, in turn, motivated certain children to focus on metaphonologic events associated with their articulatory responses. Tabletop mode may have been more successful due to its association with more frequent matched gaze, which, in turn, allowed certain children to profit maximally from the clinician's instructional models and cues.

Finally, children's responses to the fantasy element in the form of PEP the robot ranged from mild amusement to considerable absorption. Unfortunately, as described above, the power of such added cognitive-affective investment may not have been assessed adequately in the present studies due to the intelligibility limitations of the speech synthesis. Whether due to their lack of understanding of the synthetic speech or the functional inadequacy of PEP's prompt, "ask for help," the result was that the children were essentially passive participants during the computer + robot mode. The original goal was not well tested in these studies (i.e., that children would become involved with PEP as a fantasy pal who was happy when they did well, but who left the teaching to the clinician). As reported previously, the verbal behavior of some children indicating sustained interest and affection for PEP suggests that thematic, fantasy materials do have excellent potential.

CONCLUSIONS

In addition to the five conclusions about microcomputer-assisted management at the stabilization phase of response development (Shriberg et al., 1989b), the findings of this study suggest the following five conclusions about the two subphases of response evocation.

1. The clinician will likely need to be the primary facilitator for speech change during early-phase response evocation, providing moment-to-moment shifts in instructional strategies and knowledge of performance in response to fragile and transient target response attempts by the child. Even with well-developed speech recognition capabilities, the computer is incapable of anticipating, recognizing, and responding to the subtle articulatory and motivational cues from which nascent target responses emerge.

2. For some children, computer graphics can be so engaging that they compete with both the external instructional cues provided by the clinician and those internal, sensory-motor, and earliest metaphonologic events that presumably occur while learning new articulatory behaviors. Thus, for some children, computer-assisted instruction might be deferred until the target sound is under auditory control.

3. For those children for whom computer graphics do prove helpful rather than distracting at early- or later-phase response evocation, the software must provide immediate displays of stimulus and response graphics. Currently, the Apple IIe environment cannot keep pace with the type of rapid instructional cycles required during the earliest phases of response evocation. Moreover, to the extent that software requires speech synthesis, the speech cues should be maximally intelligible.

4. Microcomputers have excellent potential for engaging children in the drill activities required during later-phase response evocation. If made flexible enough to meet the learning and child factors discussed here, they offer attractive alternatives to standard clinical materials.

5. The use of fantasy and thematic materials that engage young children, challenge them, and encourage them to take active roles toward mastery and control has excellent potential and warrants programmatic research.

Most generally, the findings of considerable individual differences in the potential effectiveness and engagement value of microcomputer-assisted instructional modes is consistent with current trends in the general and special education literatures in microcomputer-assisted instruction (e.g., Carrier & Sales, 1987; Lehrer, Harckham, Archer, & Pruzek, 1986; Salisbury, 1988; Yang, 1987). Now that computers have demonstrated their potential for instructional effectiveness (cf. Niemiec & Walberg, 1987), it is time to turn from group designs that compare traditional to computer-assisted intervention toward an examination of those individual difference variables that must be accounted for in an eventually effective instructional technology.

ACKNOWLEDGMENTS

We thank Sara Hoffman, Annette Ortiz, and Carmen Rasmussen for effective, efficient, and engaged coding of the videotapes; Doris Kistler for informative statistical consultation; and Barri Babow for competent editorial assistance. We also thank the following former student clinicians for their capable participation in Study 1 and Study 2: Linda Adams, Yvonne Annis, Karen Dennis, Joanne Graziul, Sara Hoffman, Judy McCrary, Leslie Richards, Kelly Rowan, Virginia Schoeppel, Allison Sharp, and Linda Stueber. This work was supported by grants from the United States Department of Education, Research in Education of the Handicapped Program, G008400633, and the National Institute of Deafness and Other Communicative Disorders, NS-26246.

REFERENCES

BERNTHAL, J. E., & BANKSON, N. W. (1988). *Articulation and phonological disorders*. Englewood Cliffs, NJ: Prentice-Hall.

BLUME-COHEN, V. (1985). A reexamination of feedback in computer-based instruction: Implications for instructional design. *Educational Technology, 25*, 33–37.

CARRIER C. A., & SALES, G. C. (1987). A taxonomy for the design of computer-based instruction. *Educational Technology, 27*, 15–17.

COSTELLO, J. (1984). *Speech disorders in children: Recent advances*. San Diego, CA: College-Hill.

CREAGHEAD, N. A., NEWMAN, P. W., & SECORD, W. A. (1989). *Assessment and remediation of articulatory and phonological disorders* (2nd ed.). Columbus, OH: Merrill.

ECHO IIB SPEECH SYNTHESIZER [Computer peripheral]. (1986). Santa Barbara, CA: Street Electronics.

EPP, M. (1987). *VCR CONTROLLER* [Computer program]. Madison: University of Wisconsin, Waisman Center Research Computing Facility.

GROVER, S. C. (1986). A field study of the use of cognitive-developmental principles in microcomputer design for young children. *Journal of Educational Research, 79*, 325–332.

HARDCASTLE, W., JONES, W., KNIGHT, C., TRUDGEON, A., & CALDER, G. (1989). New developments in electropalatography: A state-of-the-art report. *Clinical Linguistics and Phonetics, 3*, 1–38.

HEYWOOD, G. (1986). Forms and functions of feedback in CAI and CAL. *Journal of Computer Assisted Learning, 2*, 83–92.

HOFFMAN, P. R., SCHUCKERS, G. H., & DANILOFF, R. G. (1989). *Children's phonetic disorders*. Boston: Little, Brown.

KEATING, D. L., EVANS, A. L., WYPER, D. J., & CUNNINGHAM, E. (1986). A comparison of the intelligibility of some low cost speech synthesis devices. *British Journal of Disorders of Communication, 21*, 167–172.

KELLY, L., & CHIAL, M. (1988). *A comparison of the intelligibility of five speech synthesizers*. Paper presented at the American Speech-Language-Hearing Association Annual Convention, Boston.

LEHRER, R., HARCKHAM, L. D., ARCHER, P., & PRUZEK, R. M. (1986). Microcomputer-based instruction in special education. *Journal of Educational Computing Research, 2*, 337–355.

LEPPER, M. R. (1985). Microcomputers in education: Motivational and social issues. *American Psychologist, 40*, 1–18.

MASSEY, H. J. (1988). Language-impaired children's comprehension of synthesized speech. *Language, Speech, and Hearing Services in Schools, 19*, 401–409.

MCREYNOLDS, L. V., & HUSTON, K. (1971). Token loss in speech imitation training. *Journal of Speech and Hearing Disorders, 36*, 486–495.

NIEMIEC, R., & WALBERG, H. J. (1987). Comparative effects of computer-assisted instruction: A synthesis of reviews. *Journal of Educational Computing Research, 3*, 19–37.

OLSON, D. (1987). *RELYONME: A program for estimating inter-observer reliability* [Computer program]. Madison: University of Wisconsin, Waisman Center Research Computing Facility.

SALISBURY, D. F. (1988). When is a computer better than flashcards? *Educational Technology, 28*, 26–32.

SHELTON, R., & MCREYNOLDS, L. (1979). Functional articulation disorders: Preliminaries to treatment. In N. Lass (Ed.), *Speech* *and language: Advances in basic research and practice* (Vol. 2, pp. 1–111). New York: Academic Press.

SHRIBERG, L. D. (1986). *PEPPER: Programs to examine phonetic and phonologic evaluation records* [Computer program]. Hillsdale, NJ: Lawrence Erlbaum.

SHRIBERG, L. D., & KWIATKOWSKI, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders, 47,* 256–270.

SHRIBERG, L. D., KWIATKOWSKI, J., BEST, S., HENGST, J., & TERSELIC-WEBER, B. (1986). Characteristics of children with phonologic disorders of unknown origin. *Journal of Speech and Hearing Disorders, 51,* 140–161.

SHRIBERG, L. D., KWIATKOWSKI, J., & SNYDER, T. (1986). Articulation testing by microcomputer. *Journal of Speech and Hearing Disorders, 51,* 309–321.

SHRIBERG, L. D., KWIATKOWSKI, J., & SNYDER, T. (1989a). *PEPTALK* [Computer program]. Tucson, AZ: Communication Skill Builders.

SHRIBERG, L. D., KWIATKOWSKI, J., & SNYDER, T. (1989b). Tabletop versus microcomputer-assisted speech management: Stabilization phase. *Journal of Speech and Hearing Disorders, 54,* 233–248.

SPEECHVIEWER™ [Computer program]. Boca Raton, FL: International Business Machines.

STEINBERG, E. R. (1984). *Teaching computers to teach.* Hillsdale, NJ: Lawrence Erlbaum.

STOKES, T. F., & BAER, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavioral Analysis, 10,* 349–367.

VER HOEVE, J. N. (1986). *TERMITE: A behavioral observation system for the Harris/800 minicomputer* [Computer program]. Madison: University of Wisconsin, Waisman Center Research Computing Facility.

VISI-PITCH™ [Electronic device]. Pine Brooks, NJ: Kay Elemetrics.

WATSON, C. S., REED, D. J., KEWLEY-PORT, D., & MAKI, D. (1989). The Indiana speech training aid (ISTRA) I: Comparisons between human and computer-based evaluation of speech quality. *Journal of Speech and Hearing Research, 32,* 245–251.

YANG, J.-S. (1987). Individualizing instruction through intelligent computer-assisted instruction: A perspective. *Educational Technology, 27,* 7–15.

ZION, G. (1958). *No roses for Harry.* New York: Harper & Row.

ZION, G. (1965). *Harry by the sea.* New York: Harper & Row.

Received August 28, 1989
Accepted November 6, 1989

Requests for reprints should be sent to Lawrence D. Shriberg, Ph.D., Phonology Project, Waisman Center on Mental Retardation and Human Development, 1500 Highland Avenue, Madison, WI 53705.

APPENDIX A

Clinician Protocol

Protocols for the tabletop, computer, and computer + robot modes used for Study 1 and 2 follow. Changes in the Study 1 protocols for Study 2 are indicated in bold type in parentheses.

Protocol element	*Tabletop mode*
Introduction	Seat the child at the worktable. Say: "Let's practice your ___ sound. Today we'll use the chips. **(We'll do 10 short practices.)** Every time you finish a practice I'll tell you part of a story. **(You can help me tell the story by turning the page for me when I tell you.)**" Set 10 **(6)** chips to serve as trial items in a half-circle in front of the child. The arrangement should parallel the arrangement of the trial items on the introduction-practice screen on the computer. Space the chips far enough apart so that the paper frame can fit around the chip. Set the paper frame on the first chip. Say, "I'll help you practice your sound. In this game good sounds will be like magic. Watch what happens. Say your sound the old way—no move" (hold the paper frame on the first chip). "Try to say your sound the new way—and make the box move" (move the paper frame to the second chip). "Say your sound just like me—and make the box move (move the paper frame to the third chip). If you don't make the box move the first time you can try again. I'll help you." Place the paper frame on the first chip again and say, "Every time you finish a practice—(move the paper frame from chip to chip. End by placing the paper frame below the last chip)—I will tell you part of a story." Say, "Now I'll get ready for our game." Place 10 **(5)** chips in a semicircle similar to the way they appear on the computer screen. Then place the paper frame on the first chip and say, "Time to practice your sound so you can hear part of my story."
Practice	For each of the 10 **(5)** trials in the 5 **(10)** trial blocks present the model for the child to imitate at the linguistic level that was determined from baseline testing. First announce the trial number (e.g., say "Let's do Number 1." "Now 2," or some such statement). Then present the stimulus. If the child's response is *correct*, move the paper frame to the next chip and immediately say one of the following in a random order: "That's it," "Good," "You got it." If the child's response is close, move the paper frame to the next chip and immediately say one of the following in a random order: "Nice try." "Good try." "You're trying." If the child's response is *incorrect*, do not move the paper frame to the next chip, and immediately say one of the following in a random order: "Not quite," "Not right," "Not yet." **("Oops, Let me help you.")** Repeat the trial. If this is the child's second attempt to produce the target, use the cues designated for the second attempt; if this is the third attempt, use the cues designated for the third attempt.
First four **(nine)** reinforcers	After the child completes the last trial in the first 4 **(9)** trial blocks, say, "You finished a practice. Now you can see part of my story." **("You can help me tell the story by turning the page when I tell you.")** Place the storybook in front of the child so that you cover up the chips and read part of the story. **(Tell the child "turn the page" when appropriate).**
Returning to the practice	At the end of each part of the story, say, "Time to practice so we can see more of the story." Repeat this practice to reinforcement sequence 5 **(10)** times.
Final reinforcer	After the child completes the last trial in the final block say, "You finished the last practice. Now you can see the end of my story." Read the final part of the story.
Ending the task	At the end of the final part of the story, say, "We're done practicing. You did a good job. Let's practice again on another day."
Reading the story	Just read the story. Do not make comments or show reactions. Do not sit close to the child or make any physical contact. The story needs to be read in an impersonal manner to parallel the robot's telling of the story on the computer.

Protocol element	*Computer mode*
Preparation	When the menu appears on the screen, select "Introduce, Then Start a Practice/Play Activity" and press Apple key. Select the No-Robot option and press Apple key. This will reveal the blue Introduction-practice screen. Press the Right Arrow key to advance to the second blue screen.
Introduction	Seat the child at the computer table. The blue Introduction-practice screen should be visible. Say, "Let's practice your ____ sound. Today we'll use the computer. **(We'll do 10 short practices.)** Every time you finish a practice, PEP, the robot in the computer, will tell you part of a story." **("You can help PEP tell the story by pressing this key [point to the red-dotted advance key] on the computer every time you see a blinking box at the bottom of the screen.")** Press <right arrow> to reveal the next Introduction-practice screen. Press <right arrow> again to put the flashing frame on the first trial item. From this point on, press <right arrow> whenever you finish presenting the portion of text that is written on the screen. Follow the screen text as you say, "I'll help you practice your sound. In this game good sounds are magic. Watch what happens. Say your sound the old way." Press <right arrow> "No magic." Press <right arrow>. "Try to say your sound the new way," (press <right arrow>) "and make a little magic" ("**and make magic**"). Press <right arrow> to reveal the partial display (**full display**). "Say your sound like me," (press <right arrow>) "and make magic." Press <right arrow> to reveal the full display. "If you don't make magic the first time you can try again. I'll help you." "Every time you finish a practice," (press <right arrow> to move the flashing frame through the grid of trials to the last trial item on the screen) "just like this, PEP, the robot on the computer, will tell you part of a story." (**"It will be your turn to work the computer to help PEP, the robot, tell you part of a story."**) Press <right arrow> to proceed to the next screen. Say, "I'm getting ready for our game." Press <right arrow> again to load the program. It will take a while for the program to load. During this loading period the HELP menu will be on the screen. During the waiting period, if you need to, say, "Just a little while longer" or "It's almost ready." When the program is loaded a story title will appear on the screen. Immediately press any key to reveal the empty practice screen. Then say, "Time to practice your sound so you can see part of the story," and press <right arrow> to reveal the 10 (**5**)-trial practice screen. The flashing frame will be centered on the first trial item.
Practice	For each of the trials in the trial block, present the model at the linguistic level that was determined from baseline testing for the child to imitate. First announce the trial number (e.g., say, "Let's do Number 1." "Now 2," or some such statement). Then present the stimulus. If the child's response is *correct*, press <up arrow> to reveal the full display and automatically advance the flashing frame. Immediately say the feedback phrase that appears on the screen; the phrase is one of the following in random order: "That's it," "Good," "You got it." If the child's response is *close*, press <slash> to reveal the partial display (**the full display**). Immediately say the feedback phrase that appears on the screen; the phrase is one of the following in random order: "Nice try," "Good try," "You're trying." When the flashing frame reappears on the screen, press <right arrow> to advance the frame. If the child's response is *incorrect*, press <down arrow> to reveal no visual display. Immediately say the feedback phrase that appears on the screen; the phrase is one of the following in random order: "No," Not quite," "Not right." **(Immediately say, "Oops, let me help you.")** Repeat the trial. If this is the child's second attempt to produce the target, use the cues designated for the second attempt; if this is the third attempt, use the cues designated for the third attempt.
First four (**nine**) reinforcers	After the child completes the last trial in the first 4 (**9**) trial blocks, say, "You finished a practice. Now you can see part of PEP's story." Press <right arrow> to proceed to the Story screen. **(Then say, "You can help PEP tell the story by pressing this key (point) when you see this blinking box" and point to the flashing box on the screen to the right of the text on the screen. During the story, prompt the child to press <right arrow> as needed to progress efficiently through the story.)**
Check on understanding PEP's speech	To be certain that the child understands the story, ask the child after the first screen, "Do you know what PEP said?" If the child says, "No," say, "PEP said . . ." (i.e., repeat what PEP said). Then say, "Listen again," and press <back arrow> to repeat. Continue this procedure with each screen until the child indicates that he or she can understand what PEP said. At any time during the story, if the child does not understand what PEP said, repeat this procedure.
Returning to the practice	At the end of each part of the story, the <right arrow> and <left arrow> key prompt will appear at the bottom of the story screen. Press <right arrow> and say, "Time to practice so you can see more of the story."
Final reinforcer	Repeat this practice to reinforcement sequence 5 (**10**) times. After the child completes the last trial in the last trial block, say, "You finished the last practice. Now you can see the end of PEP's story." Press <right arrow> to proceed to the Story screen.
Ending the task	At the end of the final story part the screen will freeze on the End screen and PEP will say "Good-bye," and ER (PEP's dog) will say "ɝ." Press <right arrow> to reveal the empty practice screen. Say, "We're done practicing. You did a good job. Let's practice again on another day." As you follow the text on the screen, press <right arrow> once and then again to reveal the good-bye message scrolling scross the screen. Press <right arrow> again to return to the menu.

Protocol element	*Computer + robot mode*
Preparation	When the menu appears on the screen select "Introduce, Then Start a Practice/Play Activity." Press Apple key. Select the PEP option and press Apple key.
Introduction	Seat the child at the computer table. The Menu screen, with the select-arrow pointing to the "Introduce, Then Start a Practice/Play Activity," should be visible on the screen. Say, "Let's practice your ___ sound. Today we'll use the computer. You can practice with PEP. PEP is the robot on the computer. **(You can do 10 short practices with PEP the robot on the computer.)** Every time you finish a practice with PEP, PEP will tell you part of a story." **("You can help PEP tell the story by pressing this key [point to the red-dotted advance key] on the computer every time you see a blinking box at the bottom of the screen.")** Press Apple key. The program will load, and PEP will appear on the screen. Note: From this point on you will need to press a key to continue through the program. In all cases, <right arrow> will advance the program to the next screen, and <left arrow> will allow you to repeat the current screen including the speech synthesized utterance. After PEP introduces the task and demonstrates the visual feedback system and says, "Say your sound just like your teacher and make magic," say, "If you don't make magic the first time you can try again. I'll help you." PEP will then mention the need to talk into the microphone and make reference to the reinforcer. PEP then says, "Please wait while I get ready for our game." After this utterance, press <right arrow> to load the program. It will take a while to load. During this period periodically say, "Just a little while longer," or "It's almost ready." When the program is finished loading, a story title will appear on the screen. Immediately press <right arrow> to reveal the practice screen. PEP will appear on the practice screen. After PEP says, "Time to practice your sound to see part of my story," the flashing frame will appear around the first trial item.
Practice	For each of the 10 **(5)** trials in the trial block, present the model for the child to imitate at the linguistic level that was determined from baseline testing. First announce the trial number (e.g., say, "Let's do Number 1." "Now 2," or some such statement). Then present the stimulus. If the child is *correct,* press <up arrow> to reveal the full display and automatically advance the flashing frame to the next trial item. Immediately after the full display appears, PEP will provide verbal feedback using one of the following in a random order: "That's it," "Good," "You got it." If the child is *close,* press < slash> to reveal the partial **(full)** display. Immediately after the display appears, PEP will say one of the following in a random order: "Nice try," "Good try," "You're trying." When the box cursor reappears on the screen, press <right arrow> to advance the flashing frame to the next trial item. If the child is *incorrect,* press <down arrow> to reveal no visual display, and PEP will say one of the following in a random order: "No," "Not quite," "Not right." **(PEP will say, "Oops, ask for help.")** Repeat the trial. If this is the child's second attempt to produce the target, use the cues designated for the second attempt; if this is the third attempt, use the cues designated for the third attempt.
First four **(nine)** reinforcers	After the child completes the last trial in the first 4 **(9)** trial blocks, PEP will say, "You finished a practice, now you can see part of my story." Press <right arrow> to move to the story screen. **(Then say, "You can help PEP tell the story by pressing this key [point] when you see this blinking box" [point to the box on the screen]. During the story, prompt the child to press <right arrow> as needed to progress efficiently through the story.)**
Check on understanding PEP's speech	To be certain that the child understands PEP's speech, ask the child after the first screen, "Do you know what PEP said?" If the child says, "No," say, "PEP said . . ." (i.e., repeat what PEP said). Continue this procedure with each screen until the child indicates that he or she can understand what PEP said. At any subsequent point, if the child does not understand what PEP said, repeat this procedure.
Returning to the practice	At the end of each part of the story, <right arrow> and <left arrow> keys will appear at the bottom of the screen. Press <right arrow> to return to the practice screen. PEP will say, "Time to practice so you can see more of my story." Continue to use <right arrow> to proceed through the practice. Repeat this practice to reinforcement sequence 5 **(10)** times.
Final reinforcer	After the child completes the last trial in the last trial block, PEP will say, "You finished the last practice. Now you can see the end of my story." Press <right arrow> to advance to the Story screen.
Ending the task	At the end of the final story part, PEP will say, "Good-bye." ER (PEP's dog) will say /ɝ/, and the <right arrow> and <left arrow> prompt will appear at the bottom of the screen. Press <right arrow> to advance to the practice screen. PEP will appear on the practice screen and say, "We're done practicing. You did a good job. Let's practice again on another day." Continue to press <right arrow> when directed to do so on the screen until you return to the main menu.

APPENDIX B

Efficiency, Effectiveness, Engagement Coding Protocol

The following protocol was used by the judges in Study 1 and Study 2 to identify each practice and reinforcement unit and the behaviors to be coded.

First period	*Second through next-to-last period*	*Last period*
Press 'P' immediately after the clinican says, "Let's practice your ___ sound."	Press 'P' immediately after PEP or the clinician says either, "It's time to practice" or "Are you ready to practice" (or equivalent) after the child hears a portion of the story.	Press 'P' followed by '>' immediately after PEP or the clinican says, "We're done practicing. You did a good job today. Let's practice again on another day."

Effectiveness Codes

Use of a two-letter code for each response. The first letter will code the accuracy of the child's response; the second letter will code the try.

Response type	*Behavioral signal*	*Code*
Accuracy codes		
Acceptable and correct	PEP or clinician says, "That's it," "Good," or "You got it" in response to the child's attempt to produce the target.	U
Acceptable but not correct (close)	PEP or clinician says, "Nice try," "Good try," or "You're trying" in response to the child's attempt to produce the target.	O
Unacceptable (incorrect)	PEP or clinician says, "No," "Not quite," or "Not right" in response to the child's attempt to produce the target.	I
Try codes		
First try	Child's first attempt to produce the target in the trial.	T
Second try	Child's second attempt to produce the target in the trial.	R
Third try	Child's third attempt to produce the target in the trial.	E
Additional codes		
Clinician error	Clinician behavior that cannot be reliably coded.	Z
Other responses	Includes the following: (a) child's response is premature (e.g., the child responds before the clinican presents the stimulus), (b) clinician inappropriately presents a stimulus (e.g., the child's response was close, and the clinician inappropriately presents another stimulus for the trial), (c) the clinician failed to provide knowledge of results (e.g., the clinician did not evaluate the child's response because she could not hear it or could not make a judgment and instead asks the child to repeat the try). If the clinician adds a trial to the trial block do not use the Other Response code; code the child's responses with the Accuracy and Try codes.	Y

Efficiency Codes

Period	*Behavioral signal*	*Code*
Instructions	Immediately after the clinician says, "Let's practice your ___ sound."	J
Preresponse		
First preresponse	Immediately after PEP or the clinician says, "Time to practice your sound."	K
Subsequent preresponses	Immediately after the clinician says, "Let's do Number ___" to identify the trial number or immediately after the clinician models the target, whichever comes first.	
Response	Immediately after the child's first attempt to produce the target in the trial.	L
Postresponse	Immediately after PEP or the clinician indicates that the response is acceptable (correct or close) or after the third try, whichever comes first.	;
Reinforcement	Immediately after PEP or the clinician says, "You finished a (the last) practice, now you can see part (the end) of the story (or equivalent)."	X
Interruption	Immediately when any technical interruption occurs (e.g., breakdown in equipment, interruption, biological needs such as nose blowing, etc.)	OO

Engagement Codes

General Guidelines

Eye Gaze: Judgments are made on the basis of the still-framed screen image at the end of the Judging period. The following judging sequence is to be used:

1. Look where the child (clinician) is directing the iris of the eye. If you can make a decision, assign a code.
2. If you were not able to assign a code in (1), look at the angle (up/down) and direction (right/left) of the child's (clinician's) head. If you can make a decision, assign a code.
3. If you were not able to assign a code in (2) because you were not able to judge the direction of the clinician's gaze, assign the code EL. However, if you were not able to assign a code because you were not able to judge the direction of the child's gaze, then look at the direction of the child's head and determine a 90-degree angle of vision. If that angle permits a match to the clinician's gaze, then assume a matched gaze and assign the appropriate code (EJ or EK). If that angle does not permit a match to the clinician's gaze but is directed at the materials or clinician, assign the appropriate nonmatch code (ED or EF). If that angle does not include either the clinician or the materials, then assign the code for Looking Elsewhere (EF).

Verbal: Judgments are made, as they occur, throughout the 1-s Judging period. The following judging sequence is to be used.

1. Decide if there was or was not an oral behavior. Oral behaviors are divided into two categories: Verbal and Other (which includes biologic and affective). If there is no oral behavior, assign VR.
2. Decide if the present oral behavior is verbal or not. Verbal includes true words that have a universal meaning in the language and the child's attempts to produce the speech target. Verbal does not include utterances such as "sh" to indicate "be quiet," "tsk" to indicate disapproval, colorful action terms such as "zap," cynical laughter, and so forth. Although such affective utterances do communicate the communicator's intent, they can be interpreted only after making several assumptions in the communication context. Those assumptions may not be reliable, and the resulting codes would be unreliable. If there was no verbal behavior, assign VN.
3. Assign the appropriate verbal behavior code. If a verbal code cannot be assigned, code the behavior VI.

Facial Expression: Judgments are made on the basis of the still-framed screen image at the end of the Judging period. Make judgments with the aid of a 5-point scale anchored in the following way: FF = 1–1.5 as negative; FS = 2–3.5 as neutral; FJ = 4–5 as positive.

Category	*Subcategory*	*Description of the behavior*	*Code*
Eye gaze	Matched I	Child and clinician look at each other. Eye contact is not required; what is required is that each be looking at some part of the other's body (e.g., the child may be looking at the clinician's eyes, while the clinician is looking at the child's hands).	EJ
	Matched II	Child and clinician look at computer/task materials. Clinician and child do not need to be looking at exactly the same materials. Note: if the child is looking at the monitor behind the child and clinician, that qualifies as looking at the computer unless the clinician has told the child to look at the front monitor.	EK
	Nonmatched I	Child looks at clinician while clinician looks at computer/task materials.	ES
	Nonmatched II	Child looks at computer while clinician looks at computer/task materials.	ED
	Looking elsewhere	Child looks elsewhere (i.e., not at clinician or computer/task materials). Note: if the child looks at the monitor behind the child and clinician after the clinician has told the child to look at the front monitor, code the behavior EF.	EF
	Cannot judge	Cannot determine direction of clinician's gaze while child is looking at clinician or computer/task materials or cannot determine direction of the child's gaze.	EL
Verbal	Neutral or positive	Includes (a) task-related responses (i.e., attempts to produce the stimulus), (b) positive or neutral statements about the materials or task (e.g., positive: "I like this story"; neutral: "Are you going to use all the chips?").	VJ
	Negative	Includes only negative comments (e.g., "I don't want to do this any more," "I hate this.").	VF
	Other	Includes all comments that are not task related. These comments can be positive, neutral, or negative.	VU
	Cannot judge	Includes only responses that cannot be judged.	VI
	Other oral	Includes oral behaviors that are not verbal, including biologic behaviors (e.g., sneeze, cough, hiccup) and affective behaviors (e.g., crying, laughing, snorting, whining, nonspeech vocalizations, yelled nonspeech vocalizations).	VN
	No oral	Includes the absence of all oral behavior including verbal, biologic, and affective.	VR
Facial expression	Smile	The corners of the child's mouth are curved upward (into a smile).	FJ
	Frown	The child's forehead is wrinkled or contracted (into a scowl) and lips are protruded (into a pout) or the corners of the mouth are curved down (into a frown), or the child appears puzzled (wrinkles in forehead and other parts of face suggest that the child may not understand something).	FF
	Neutral	Includes all facial expressions that are neither FJ or FF (as defined by the five-point scale guide).	FS
	Cannot judge	Includes expressions that cannot be judged because (a) the child's face cannot be seen because the clinician is touching (at a minimum) a portion of the child's lips, (b) the clinician dictated the child's facial posture because the clinician's hands were on the child's face in such a way as to constrain the muscles used in facial expression, (c) the child was producing the target sound at the moment of judging.	FI