

Improving the Vocabulary of Children with Hearing Loss

Dominic W. Massaro, Ph.D., and Joanna Light, M.S.

The goal of this study was to test the effectiveness of a Language Wizard/Player with Baldi, a computer-animated tutor, for teaching new vocabulary items to children with a hearing loss. Eight students with hearing loss, between the ages of 6 and 10, were tested and trained for about 20-30 minutes a day, 2 days a week for about 10 weeks on three categories of eight words each. The design of the experiment was based on a within-student multiple baseline design in which all three categories of words were continuously being tested while one of the categories was being trained. Knowledge of the words remained negligible without training and learning occurred fairly quickly for all words once training began, reaching asymptotic levels in each category. Knowledge of the trained words did not degrade after training once these words ended and training on other words took place. Finally, retention was nearly perfect, as indicated by a reassessment test 4 weeks after the experiment.

Introduction

The purpose of this study was to test the effectiveness of a Language Wizard/Player with Baldi, a computer-animated tutor (Bosseler & Massaro, 2003; Massaro, 2002) for teaching new vocabulary items to children with hearing loss. It is well known that children with hearing loss have significant deficits in both spoken and written vocabulary knowledge (Breslaw, Griffiths, Wood, & Howarth, 1981; Holt, Traxler, & Allen, 1997). One reason is that these children tend not to overhear other conversations because of their limited hearing and are thus shut off from an opportunity to learn vocabulary. The children with hearing loss often do not have names for specific things and

Dominic W. Massaro, Ph.D., is a professor of psychology and computer engineering, chair of digital arts/new media and director of the Perceptual Science Laboratory at the University of California, Santa Cruz. His research uses formal experimental and theoretical approach to the study of speech perception, reading, and psycholinguistics. He has developed an embodied conversational agent, Baldi, for spoken language synthesis, language tutoring, and "edutainment." He is co-founder of Animated Speech Corporation, a company dedicated to improving language learning for children with language challenges. Joanna Light, M.S., is currently completing an M.A. in speech and language pathology at San Jose State University. She has a B.A. in psychology and linguistics from McGill University in Montreal, Canada and an M.S. in cognitive psychology from the University of California, Santa Cruz. Her research interests are atypical speech and language development.

concepts and therefore communicate with phrases such as "the window in the front of the car," "the big shelf where the sink is," or "the step by the street" rather than "windshield," "counter," or "curb" (Barker, 2003). We begin with a review of research that highlights the importance of vocabulary knowledge for all children and the need for its direct instruction followed by a description of related studies.

Essential Role for Vocabulary Knowledge in Language Development

Although there is no consensus on the best way to teach or to learn language, there are important areas of agreement. One is the central importance of vocabulary knowledge for understanding the world and for language competence in both spoken language and in reading (Gupta & MacWhinney, 1997). Empirical evidence indicates that very young normally-developing children more easily form conceptual categories when category labels are available than when they are not (Waxman, 2002). Once the child knows about 150 words, there is a sudden increase in the rate at which new words are learned and the emergence of grammatical skill (Marchman & Bates, 1994). Even children experiencing language delays because of specific language impairment benefit once this level of word knowledge is obtained. Vocabulary knowledge is positively correlated with both listening and reading comprehension (Anderson & Freebody, 1981; Stanovich, 1986; Wood, 2001), and predicts overall success in school (Vermeer, 2001). It follows that increasing the pervasiveness and effectiveness of vocabulary learning offers a promising opportunity for improving conceptual knowledge and language competence for all individuals, whether or not they are disadvantaged because of sensory limitations, learning disabilities, or social condition.

Validity of the Direct Learning of Vocabulary

There are important reasons to justify the need for direct teaching of vocabulary. Although there is little emphasis on the acquisition of vocabulary in typical school curricula, research demonstrates that some direct teaching of vocabulary is essential for appropriate language development in children who are developing normally (Beck, McKeown, & Kucan, 2002). Contrary to a common belief that learning vocabulary is a necessary outcome of reading in which new words are experienced in a meaningful context, context seldom disambiguates the meaning of a word completely. As an example, consider the passage from *The Fir Tree* by Hans Christian Andersen:

Then two servants came in rich livery and carried the fir tree into a large and splendid drawing-room. Portraits were hanging on the walls, and near the white porcelain stove stood two large Chinese vases with lions on the covers.

Most of the words are not disambiguated by context. The meaning of *livery*, *portraits*, *porcelain*, and *vases*, for example, cannot be determined from the context of the story alone. Research by Beck et al. (2002) and Baker, Simmons, and Kameenui (1995) provides some evidence that children with normal hearing more easily acquire new vocabulary by direct intentional instruction than by other incidental means. Although we are unable to find relevant research, we would expect that the same advantage of direct instruction would exist for children with hearing loss.

Knowing a word is not an all-or-none proposition. A single experience with a word (even if the correct meaning of the word is comprehended) is seldom sufficient for mastering that word. Acquiring semantic representations appears to be a gradual process that can extend across several years (McGregor, Friedman, Reilly, & Newman, 2002). The completeness of these semantic representations will, therefore, vary. Words are complex multidimensional stimuli, and a person's knowledge of the word will not be as complete or as accurate as its dictionary entry. Semantic naming errors are more likely to occur with those items that have less embellished representations. Thus, it is important to overtrain or continue vocabulary training after the word is apparently known, and to present the items in a variety of contexts in order to develop rich representations. Picture naming and picture drawing are techniques that can be used to probe and reinforce these representations (McGregor et al, 2002). Qian (2002) found that the dimension of vocabulary depth (as measured by synonymy, polysemy, and collocation) is as important as that of vocabulary size in predicting performance on academic reading. Therefore, a student can profit from the repeated experience of practicing new words in multiple contexts during the direct teaching of vocabulary.

Effectiveness of Computer-Based Instruction

Computer-based instruction is an emerging method to train and develop vocabulary knowledge for both native and second-language learners (Druin & Hendler, 2000; Wood, 2001) and individuals with special needs (Moore & Calvert, 2000; Barker, 2003; Heimann, Nelson, Tjus, & Gilberg, 1995). An incentive to employing computer-controlled applications for training is the ease with which automated practice, feedback, and branching can be programmed. Another valuable component of computer-based instruction is the potential to present multiple sources of information, such as text, sound, and images in parallel (Dubois & Vial, 2000; Chun & Plass, 1996). Incorporating text and visual images of the vocabulary to be learned along with the actual definitions and spoken words facilitates learning and improves memory for the target vocabulary. Dubois and Vial (2000), for example, found an increase in recall of second-language vocabulary when training consisted of combined presentations of spoken words, images, written words, and text relative to only a subset of these formats.

Baldi®: Visible Speech, Realism, and Student Engagement

Computer-based instruction makes it possible to include embodied conversational agents rather than simply text or disembodied voices in lessons. Baldi® is a computer-animated agent who provides accurate visible and audible speech in the tutoring situation. There are several reasons why the use of auditory and visual information from a talking head like Baldi is so successful, and why it holds so much promise for language tutoring (Massaro, 1998). These include a) the information value of visible speech, b) the robustness of visual speech, c) the complementarities of auditory and visual speech, and d) the optimal integration of these two sources of information.

The face presents visual information during speech that is critically important for effective communication. While the auditory signal alone is adequate for communication, visual information from movements of the lips, tongue, and jaws enhance intelligibility of the acoustic stimulus (particularly in noisy environments). Adding visible speech can often double the number of recognized words from a degraded auditory message. In a series of experiments, we asked college students with normal hearing to report the words of short sentences (Jesse, Vrignaud, & Massaro, 2002). These sentences were presented in noise in order to produce some errors. On some of the trials, only the noisy auditory sentence was presented. On other trials, the noisy auditory sentence was aligned with Baldi, our synthetic talking head. The presence of Baldi facilitated performance for each of the 71 participants. Performance was more than doubled for those participants performing relatively poorly given auditory speech alone. Moreover, speech is enriched by the facial expressions, emotions, and gestures produced by a speaker (Massaro, 1998).

The visual components of speech offer a lifeline to those with severe or profound hearing loss. Even for individuals who hear well, these visible aspects of speech are especially important in noisy environments. For individuals with severe or profound hearing loss, understanding visible speech can make the difference between effectively communicating orally with others and a life of relative isolation from oral society (Trychin, 1997). Erber (1972) tested three populations of children (adolescents and young teenagers): normal hearing (NH), severely impaired (SI), and profoundly deaf (PD). The test consisted of a videotaped speaker pronouncing the eight consonants /b, d, g, p, t, k, m, n/ spoken in a bi-syllabic context /aCa/, where C refers to one of the eight consonants. Although all three groups benefited from seeing the face of the speaker, and the group revealed with severe impairment had the largest performance gain in the bimodal condition relative to either of the unimodal conditions (Massaro & Cohen, 1999). The group with normal hearing had very good auditory information so that the face could not contribute much whereas the group who was profoundly deaf had very poor auditory information so that the voice didn't contribute as much. The group with severe impairment, on the other hand, had a reasonable degree of both auditory and visual information. As noted in the following discussion of complementary

and optimal integration, perception of speech can be very good when some hearing is present and the face of the speaker can be seen.

Empirical findings indicate that speechreading, or the ability to obtain speech information from the face, is robust; that is, perceivers are fairly good at speechreading in a broad range of viewing conditions. To obtain information from the face, the perceiver does not have to fixate directly on the talker's lips but can be looking at other parts of the face or even somewhat away from the face. Furthermore, accuracy is not dramatically reduced when the facial image is blurred (because of poor vision, for example), when the face is viewed from above, below, or in profile, or when there is a large distance between the talker and the viewer (Jordan & Sergeant, 2000; Massaro, 1998; Munhall & Vatikiotis-Bateson, 2004). These findings indicate that speechreading is highly functional in a variety of non-optimal situations. The robustness of the influence of visible speech is illustrated by the fact that people naturally integrate visible speech with audible speech even when the temporal occurrence of the two sources is displaced by about a 1/5 of a second (Massaro & Cohen, 1993). Given that light and sound travel at different speeds and that the dynamics of their corresponding sensory systems also differ, a crossmodal integration appears to be relatively immune to small temporal asynchronies (Massaro, 1998).

A visual talking head allows for complementarities of auditory and visual information. Auditory and visual information are complementary when one of these sources is most informative in those cases in which the other is weakest. Because of this, a speech distinction between segments is differentially supported by the two sources of information. That is, two segments that are robustly conveyed in one modality are relatively ambiguous in the other modality (Massaro & Cohen, 1999). For example, the difference between /ba/ and /da/ is easy to see but relatively difficult to hear. On the other hand, the difference between /ba/ and /pa/ is relatively easy to hear but very difficult to discriminate visually. The fact that two sources of information are complementary makes their combined use much more informative than would be the case if the two sources were non-complementary or redundant (Massaro, 1998).

The final value afforded by a visual talking head is that perceivers combine or integrate the auditory and visual sources of information in an optimally efficient manner (Massaro, 1987; Massaro & Cohen, 1999; Massaro & Stork, 1998). There are many possible ways to treat two sources of information: use only the most informative source, average the two sources together, or integrate them in such a fashion that both sources are used but that the least ambiguous source has the most influence. Perceivers in fact integrate the information available from each modality to perform as efficiently as possible (Massaro, 1998).

One might question why perceivers integrate several sources of information when just one of them might be sufficient. Most of us do reasonably well in communicating over the telephone, for example. Part of the answer might be

grounded in our ontogeny. Integration might be so natural for adults even when information from just one sense would be sufficient because, during development, there was much less information from each sense and therefore integration was all the more critical for accurate performance (Lewkowicz, 2004).

Baldi, our 3-D computer-animated talking head, provides realistic visible speech that is almost as accurate as a natural speaker (Cohen, Beskow, & Massaro, 1996; Massaro, 1998). The quality and intelligibility of Baldi's visible speech has been repeatedly modified and evaluated to accurately simulate a naturally talking human (Massaro, 1998). Baldi's visible speech can be appropriately aligned with either synthesized or natural auditory speech. Baldi also has teeth, a tongue, and a palate to simulate the inside of the mouth, and the tongue movements have been trained to mimic natural tongue movements (Cohen, Walker, & Massaro, 1998). We have also witnessed that the student's engagement is enhanced by face-to-face interaction with Baldi (Bosseler & Massaro, 2003; Massaro & Light, 2004). In this study, we test the hypothesis that this technology has the potential to help individuals with hearing loss learn vocabulary.

The Language Wizard/Player with Baldi

Our Language Wizard/Player with Baldi is a user-friendly platform for creating and presenting language lessons. Guided by a user-friendly Wizard, coaches create lessons by first inputting a set of images representing the vocabulary to be learned. Each image is associated with a word or description. There are a number of optional exercises that allow the student to be tested and tutored by Baldi on these items. The tutoring includes the association of the images with their spoken and written referents, speaking the items, and reading and spelling the items. Within the Player, Baldi guides the student through the testing and tutoring, and gives feedback and encouragement. All of the results of each lesson are recorded for later analyses. A more detailed description of this learning platform is given in the Method Section.

Relationship of Language Wizard/Player to Language Learning

The Language Wizard/Player with Baldi encompasses and instantiates the developments in the pedagogy of how language is learned, remembered, and used. Research in education has shown that children with normal hearing can be taught new word meanings by using drill and practice methods (e.g., McKeown, Beck, Omanson, & Pople, 1985; Pany & Jenkins, 1978; Stahl, 1983). It has also been demonstrated that direct teaching of vocabulary by computer software is possible and that an interactive multimedia environment is ideally suited for this learning (Wood, 2001). Wood (2001) observes,

Products that emphasize multimodal learning, often by combining many of the features discussed above, perhaps make the greatest contribution to dynamic

vocabulary learning. Multimodal features not only help keep children actively engaged in their own learning, but also accommodate a range of learning styles by offering several entry points. When children can see new words in context, hear them pronounced, type them into a journal, and cut and paste an accompanying illustration (or create their own), the potential for learning can be dramatically increased.

Following this model, many aspects of our lessons enhance and reinforce learning. For example, the existing Language Player (Bosseler & Massaro, 2003) makes it possible for the student to

- Observe the words being spoken by a realistic talking interlocutor (Baldi),
- See the word written as well as spoken,
- See visual images of referents for the words,
- Click on or point to the referent,
- Hear himself or herself say the word,
- Read and spell the word by typing, and
- Observe the word used in context.

Additional Benefits

Other benefits of our program include the ability to seamlessly meld spoken and written language, provide a semblance of a game-playing experience while actually learning, and to lead the child along a growth path that always bridges his or her current “zone of proximal development” (Vygotsky, 1986). The Wizard exploits this zone with individualized lessons, and with lessons that can bypass repetitive training when student responses indicate that material is mastered.

The Language Player provides a learning platform that allows optimal conditions for learning and the engagement of fundamental psychological processes such as working memory, the phonological loop, and the visual-spatial scratchpad (Atkins & Baddeley, 1998). Evidence by Baddeley and colleagues (Baddeley, Gathercole, & Papagno., 1998; Evans et al., 2000) supports our strategy of centering vocabulary learning in spoken language dialogs. There is also some evidence that reading aloud activates brain regions that are not activated by reading silently (Berninger & Richards, 2002). Thus, the imitation and elicitation activities in the Language Player should reinforce learning of vocabulary and grammar.

Abstract Words

A potential criticism of our approach is that many words and concepts cannot be learned via a multimedia Language Wizard/Player. However, there is evidence that thought is couched in modality-specific experiences and representations (for an excellent review, see Prinz, 2002). Perceptual processes, such

as those involved in understanding spoken language and perceiving pictures, are intricately involved in performing conceptual tasks. Learning and using a concept necessarily involves some type of connection to the senses. Namy and Gentner (2002) have proposed a model for concept acquisition in which children learn conceptual properties in addition to peripheral properties. Children benefit from having two instances of a category rather than just one, which allows an alignment process in which they discern conceptual properties over and beyond perceptual properties. Experiments usually show that perceptual properties override abstract conceptual properties, but it is only reasonable that both types of properties would be important for cognitive/linguistic representations. Thus, we believe the vocabulary and language-learning paradigm can be productively extended to include words and concepts that are usually considered to be relatively abstract. We have been successful in teaching a variety of concepts, including spatial location, singular versus plural, and actor versus recipient (Bosseler & Massaro, 2003).

Previous Research on the Educational Impact of Animated Tutors

The Language Wizard/Player with Baldi has been in use at the Tucker Maxon Oral School in Portland, OR. Children with hearing loss tend to have major difficulties in acquiring language, and they serve as particularly challenging tests for the effectiveness of our pedagogy. Barker (2003) examined if training with the animated tutor software would result in vocabulary acquisition and retention. Students were given cameras to photograph objects and surroundings at home. The pictures of these objects were then incorporated as vocabulary items in the lessons. A given lesson had between 10 and 15 items. Students worked on learning the items for about 10 minutes a day until they reached 100% on the post-test. They then moved on to another lesson. About 1 month after each successful (100%) post-test, they were re-tested on the same items. Ten girls and nine boys participated in the applications. There were six children with hearing loss and one child with normal hearing between 8 and 10 years of age in the "lower school." Ten children with hearing loss and two children with normal hearing, between ages 11 and 14, participated from the "upper school."

Given that similar results occurred for the two groups, Figure 1 gives the average results of these lessons across the two groups of children. The results are given for three stages of the study: pre-test, post-test, and retention after 30 days. The items were classified as known, not known, and learned. Known items are those that the children already knew on the initial pre-test before the first lesson. Not known items are those that the children did not know, as evidenced by their inability to identify these items in the initial pre-test. Learned items are those that the children identified incorrectly on the initial pre-test and correctly in the post-test. Students knew about half of the items without any learning, they successfully learned the other half of the items, and retained about half of the newly learned items when retested 30 days later.

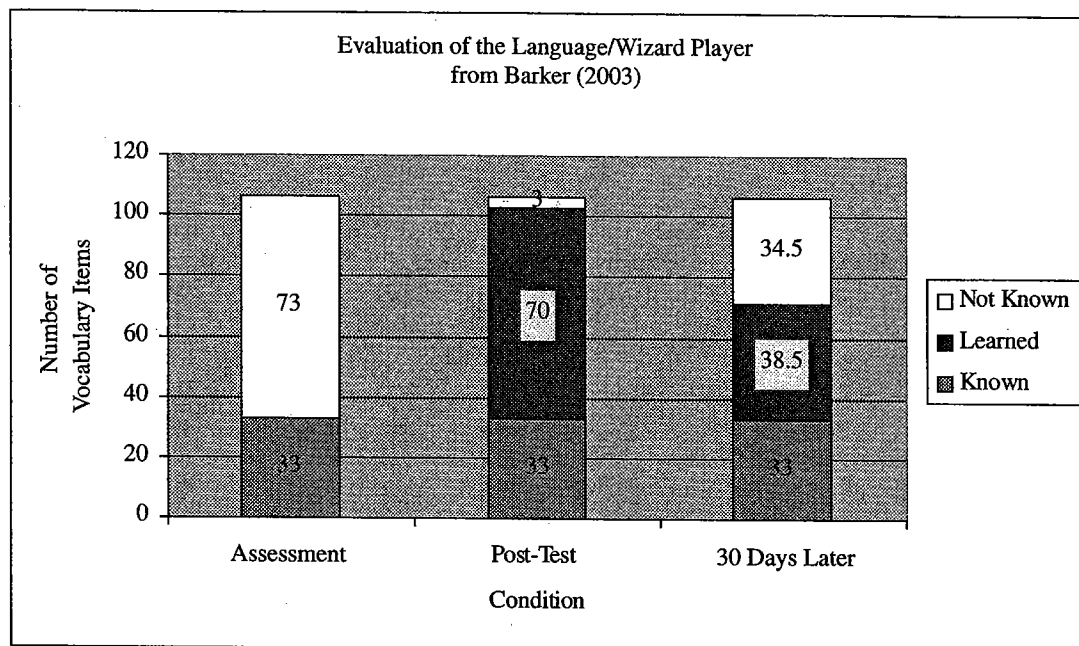


Figure 1. The average number of words that were already known, the average number learned using the program, the average number retained after 30 days, and the total amount of time spent in training. The results showed significant vocabulary learning, with about 55% retention of new words after 30 days (from Barker, 2003).

The results of the Barker (2003) evaluation in Figure 1 show that the children learned a statistically significant number of new words and retained about half of them a month after training ended. No control groups were used in that evaluation, however, and it is possible the children were learning the words outside of the tutoring environment. For example, the children could have learned the words at home or from their friends. Furthermore, the time course of learning with the vocabulary player was not evaluated. It is of interest to know how quickly words can be learned in order to give some idea of how this learning environment would compare to other situations such as the typical classroom. Finally, both identification and production of the words was assessed in the current study whereas only identification was measured previously.

Method

Eight students with hearing loss were tested and trained for about 20–30 minutes a day, 2 days a week for about 10 weeks on three categories of eight words each. The design of the experiment was based on a within-student multiple baseline design (Baer, Wolf, & Risley, 1968; Horner & Baer, 1978) where certain words are continuously being tested while other words are being tested and trained. Although the student's instructors and speech therapists agreed not to teach or use these words during our investigation, it is still possible that the words could be learned outside of the Language Player

Table 1. Age of the participants at the midpoint of the study as well as individual and average aided auditory device thresholds (dB HL) at 4 frequencies for the 8 students used in the current study. The Participants 1 and 2 were in Grade 1 and the others were in Grade 4. Participant 7 had a cochlear implant and the seven other children had binaural hearing aids except for Participant 8 who had just one aid. The Participant numbers (S#) correspond to those in the results; PTA is pure tone average; ULE and URE are unaided thresholds for left and right ears, respectively.

S#	Age	500 Hz	1000 Hz	2000 Hz	4000 Hz	PTA	ULE	URE
1	7.2	40	35	47	55	41	78	80
2	6.11	35	30	35	43	33	80	85
3	10.7	25	35	40	45	33	35	35
4	9.3	30	33	45	68	36	95	42
5	11.0	40	35	40	50	38	*	*
6	10.0	50	52	55	60	52	95	95
7	9.4	25	15	25	35	21	90	80
8	9.11	30	35	40	60	35	110	60
M		34	34	41	52	36	83	68

environment. The single student multiple baseline design monitors this possibility by providing a continuous measure of the knowledge of words that are not being trained, as well as those being trained. Thus, any significant differences in performance on the trained words and untrained words can be attributed to the Language Player training program itself rather than some other factor.

Students

Eight children with hearing loss, 2 males ages 6 and 7, and six females ages 9 and 10, were recruited from The Jackson Hearing Center (a special day school for the deaf, which is mainstreamed into Fairmeadow Elementary School) in Los Altos, CA. Parental consent was obtained to have the children participate in our study. All children were mainstreamed for certain subjects, but were in a special day class for Language Arts. The male students were in Grade 1 and the female students in Grade 4, and all students needed help with their vocabulary building skills as suggested by their special day teachers. As can be seen in Table 1, one child had a cochlear implant and the seven other children had binaural hearing aids except for one child with one hearing aid. Table 1 also gives the individual and average aided auditory thresholds for each participant.

Items

The experimenter (JL) developed a collection of vocabulary items that was individually tailored for each student in order to suit his or her vocabulary building needs as suggested by his or her special day teacher. Each collection

Table II. The categories and items of the three sets of words that were tested and trained for each participant.

<i>Participant</i>	<i>Set 1</i>	<i>Set 2</i>	<i>Set 3</i>
1	Fruits & vegetables —Cabbage, yam, mango, olives, radish, zucchini, beet, asparagus	Transportation devices —Blimp, kayak, tractor, trolley, jet-ski, stroller, yacht, unicycle	Body parts —ankle, armpit, calf, thigh, wrist, waist, chin, palm
2	Shapes —clover, cone, cylinder, oval, octagon, pentagon, pyramid, sphere	Fruits & vegetables —artichoke, leek, cabbage, beet, coconut, papaya, parsnip, persimmon	Animals —antelope, armadillo, caribou, cheetah, coyote, hyena, panther, platypus
3	Animals —armadillo, iguana, moose, panther, pelican, antelope, koala, ostrich	Transportation devices —blimp, tractor, trolley, unicycle, parachute, sailboat, yacht, jet-ski	Body parts —calf, kidney, thigh, stomach, liver, ankle, intestines, knuckle
4	Musical instruments —accordion, banjo, cymbals, fiddle, harmonica, oboe, trombone, flute	Animals —antelope, hyena, iguana, ostrich, anteater, panther, scorpion, pelican	Transportation devices —blimp, canoe, yacht, kayak, sailboat, submarine, tractor, jet-ski
5	Musical instruments —accordion, cello, cymbals, harmonica, harp, mandolin, oboe, tambourine	Fruits & vegetables —asparagus, cabbage, avocado, olives, squash, yam, zucchini, parsnip	Transportation devices —blimp, kayak, parachute, trolley, tractor, jet-ski, yacht, jet
6	Animals —anteater, iguana, coyote, panther, pelican, armadillo, antelope, ox	Musical instruments —accordion, banjo, cello, cymbals, harp, oboe, mandolin, piccolo	Transportation devices —canoe, kayak, sailboat, parachute, trolley, blimp, stroller, unicycle
7	Musical instruments —accordion, banjo, cymbals, fiddle, oboe, piccolo, trombone, tuba	Sporting equipment —barbell, cleat, club, cue, dart, oar, snorkel, tee	Transportation devices —blimp, kayak, tractor, trolley, yacht, jet-ski, raft, unicycle
8	Fruits & vegetables —artichoke, avocado, beet, eggplant, leek, mango, parsnip, yam	Musical instruments —accordion, banjo, cello, harmonica, oboe, piccolo, trombone, xylophone	Animals —antelope, coyote, pelican, koala, panther, platypus, caribou, cheetah

of items was comprised of 24 items, broken down into 3 categories of 8 items each. Table 2 lists the categories and items used for each of the students.

The experimenter used the Language Wizard to make the assessment test and the lessons that associated the visual images to spoken and written words. The Wizard was equipped with default settings that could be modified to specify what Baldi said and how he said it (e.g. by accurately specifying the rate and speed at which he spoke), the feedback given for responses,

the number of attempts permitted for the student per question, and the number of times each item was presented. The items were randomized and practiced once in each of the learning exercises. The items were randomized and tested twice in the assessment tests.

Procedure

Testing and training were carried out individually, in a quiet room at the Jackson Hearing Center, for about 20–30 minutes each day for 2 days a week for approximately 10 weeks. All children wore their personal auditory devices while participating in the study. Lessons were presented at a personal desk equipped with a laptop computer, external speakers (Amplified Sony, Model PCVA-SP1), and an external microphone (Quickshot, Model QS-5841). During the first day of the study, the intensity of the speech was set at a comfortable loudness (68.9 dB-A fast, B & K 2203 sound level meter), and was kept constant at this level throughout the study. The sound card was a Maestro Wave/WaveTable Synthesis Device provided by ESS Technology, Inc. The sampling rate for digitizing the participants' productions for playback was 8 KHz.

Images of the vocabulary items were presented on the screen next to Baldi as he spoke, as illustrated in Figure 2. The figure gives a typical lesson screen in which the vocabulary consisted of fruits and vegetables. The yellow outlined region around the zucchini represents the item being tested and/or learned. The happy and sad faces in the bottom left hand corner represent feedback for correct or incorrect responses, respectively. Some of the exercises required the child to respond to Baldi's instructions such as "click on the cabbage," or "show me the yam," by clicking on the highlighted area or by moving the computer mouse over the appropriate image until an item was highlighted and then clicking on it. Two other exercises asked the child to recognize the written word and to type the word, respectively. The production exercises asked the child to repeat after Baldi once he named the highlighted image or to name the highlighted image on their own.

Pre-testing

Prior to testing and training, a series of pilot tests was carried out individually for each student to determine the three sets of words to be used for that student. The pilot tests were the same format as the assessments, consisting of both an identification exercise and a production exercise. Three categories of 10 words each were initially composed for each student in order to find eight items in each category that were unknown to the student. Word lists were generated by the second author based on word categories as suggested by each student's special day teacher. Word lists were revised with the teacher until each list seemed appropriate for each student. Words known by

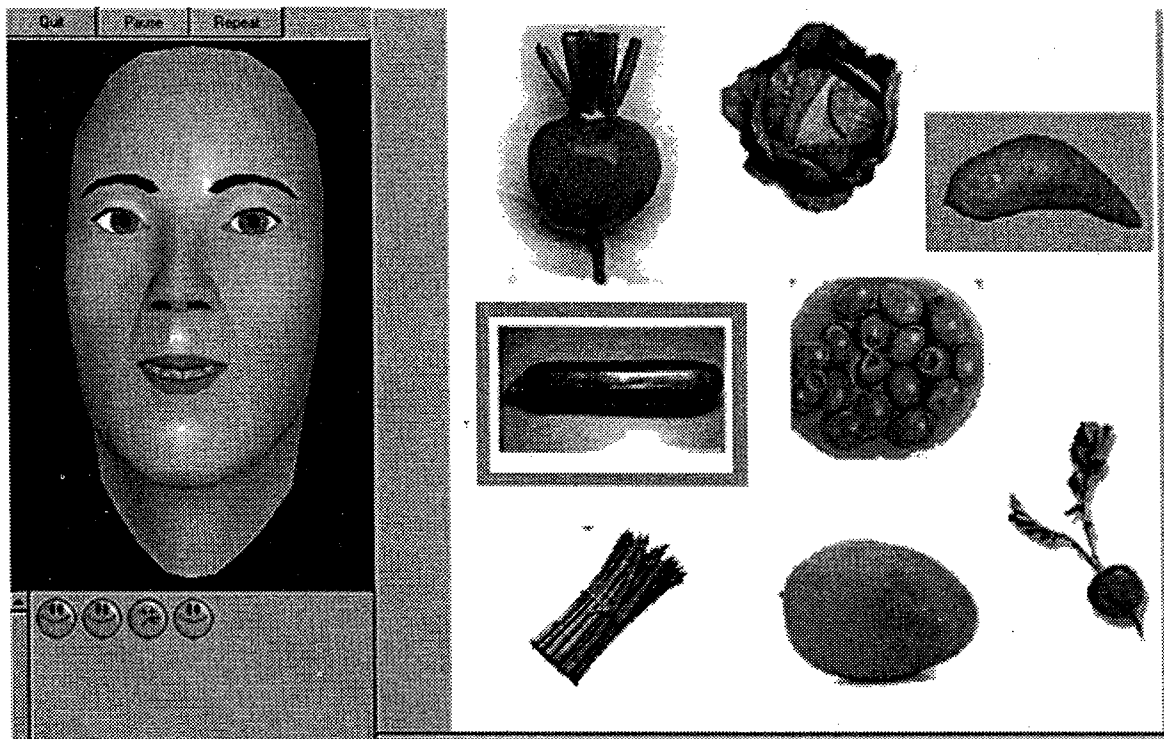


Figure 2. A screen from a language lesson on fruits and vegetables, illustrating the format of the tutors. The screen shows Baldi, the vocabulary items, and “stickers.” In this application the students learn to identify fruits and vegetables. For example, Baldi says “Click on the cabbage.” The student clicks on the appropriate region and visual feedback in the form of stickers (the happy and unhappy faces) are given for each response. When the student drags the mouse over a specific item, the region becomes highlighted, indicating the student’s selection.

the student were removed and replaced with new words. The final lists of three sets of eight words each were finalized when 3 days of pilot tests showed that these words were unknown to the student. Once the three sets of words were finalized, pre-testing continued on all three sets of words and training began on the first set.

Training

The assessment testing, described in Table 3, always occurred at the beginning of each day of training. Each of the three word sets was tested each day and the order of presentation changed in a systematic pattern from day to day (set 123, 231, 312, 123, . . .). In the identification assessment test, the child was requested to click on the image corresponding to each word. In the production assessment test, one of the images was highlighted and the child was asked to produce the vocabulary item. No feedback was given in these tests. Training on the appropriate word set followed this testing. During training, there were seven exercise modules: presentation, perception, reading,

Table III. Description of the assessment test, which was given for all three sets of words at the start of each day.

<i>Days</i>	<i>Application Module</i>	<i>Categories Involved</i>	<i>Description</i>
All days	Assessment	All Three	<p>The main performance test given at the start of each day on all three sets of words. This module was composed of both identification and production tests. No feedback was given.</p> <p>Identification: Baldi gave an instruction (e.g. "click on the (word)," "show me the (word)") and the student was required to drag the computer mouse over the correct image until it was highlighted and click on it. Items were randomly presented and each item was presented twice.</p> <p>Production: Baldi asked a question (e.g. "What is the name of this?" or "What do you call this?") and the student was required to respond after the beep by speaking into the microphone. Items were randomly presented and each item was presented twice.</p>

spelling, imitation, elicitation, and post-test (see Table 4). After the study had started, the spelling exercise was eliminated for the two Grade 1 students because it was too difficult and time consuming. The program stored each student's individual performance in a log file.

As described in Table 4, each training lesson was designed to have each student progress through a series of seven exercises: spoken presentation with pictures and written labels of each vocabulary item being trained; perception in which the student was asked to click on the image of a word and given feedback about his or her response as well as indication about the correct response; reading requiring the student to recognize the written word; spelling where the student was asked to spell the word; imitation requiring that the student repeat the words after Baldi has said them (the student's voice was recorded and played); elicitation in which the student would name the highlighted item, followed by Baldi naming aloud the item; and the post-test where the student would have to identify the highlighted item with feedback. The order of presentation of the items was always randomized and each item was presented twice in the pre-test and the post-test and once in the other exercises. In these training exercises, however, the correct item was reinforced if the student did not perform correctly on that item. For example, in the presentation exercise, Baldi might say, "This is the panther. Click on the panther." If the student clicked on another animal, Baldi might say, "Good try. This is the panther,"

Table IV. Description of the seven exercises involved in training. All exercises were given on each day of training for the set of words being learned.

<i>Application Exercise</i>	<i>Description</i>
Presentation	One image would become highlighted and Baldi would tell the student "this is a zucchini" (for example) while the written label of the vocabulary item appeared on the screen below the canvas of images. Baldi then instructed the student to "show me the zucchini" and the student was required to drag the computer mouse over the highlighted image and click on it. This was to reinforce that the student knew which image was being described. The items were randomly presented and each item was presented once.
Perception	Baldi instructed the student to "click on the zucchini" (for example) and the student was required to drag the computer mouse over the item that was just presented and click on it. Feedback was given via a happy or sad face. If the student chose the wrong item, the correct item was highlighted and Baldi told the student that the word they chose was not the zucchini and that the item that is highlighted is the zucchini. Items were randomly presented and each item was presented once.
Reading	The written text of all of the vocabulary items was presented below the images, as shown in Figure 2. Baldi instructed the student to click on the word corresponding to the highlighted image. Feedback was given. Items were randomly presented and each item was presented once.
Spelling	One of the images was highlighted while Baldi asked the student to type the corresponding word. Feedback was given. If the student was incorrect, the correct spelling of the vocabulary item appeared above the student's attempt and Baldi read the word and spelled it out to the student.
Imitation	One of the images was highlighted and Baldi named the item. The student was instructed to repeat what Baldi had just said after the tone. Once the student said the word, his or her voice was recorded and played back to him or her.
Elicitation	One of the images was highlighted and Baldi asked the student to name it. After 3 seconds, and independently of the student's production response (correct or incorrect), Baldi said the correct label.
Post-test	Baldi instructed the student to "click on the zucchini" and the student was required to drag the computer mouse over the item that was just presented and click on it. Feedback was given about the student's response via a happy or sad face. Items were randomly presented and each item was presented once.

and the correct item would be highlighted. The lesson would continue with another item regardless of the actual response of the student.

Each student was administered the training lessons in a progressive fashion until the learning criterion was obtained. The criterion was met when no

more than 1 out of 16 items (there were two identifications/productions of each item) was identified or produced incorrectly on the assessment test for 3 consecutive days. Once this criterion for one set of words was reached, testing continued on this word set, and training began on the next set of words. After the student reached the learning criterion on the third word set, the training ended. The second author was present during all training sessions and independently scored the production responses as they were produced. A correct production response was recorded when the student pronounced a word label that was comprehensible and matched the label of the vocabulary item that was being trained. An incorrect response was recorded when the student produced an incomprehensible label, a non-matching label, or no response at all. For example, if the vocabulary item being trained was "tee," "golf thing" would be scored as incorrect.

Post-testing

Assessment tests on a set of words continued after training was complete. These assessments were the same as those before and during training, consisting of both identification production. Once training was completed for the third set of words, three additional days of assessment tests were given. Student 2 was only able to complete 2 days of this assessment because of school holidays. Finally, the students were retested with the same assessment test on a single day about 4 weeks after the end of the experiment.

Results

Identification is a measure of receptive language, and production is a measure of productive language. To measure both of these behaviors, we tested the children on both identification and production of the vocabulary items. In identification, the participants were required to click on the image that corresponded to the item that Baldi asked them to find. In production, they had to orally respond with an answer after Baldi asked them to name a specific item that was highlighted on the computer monitor. Figures 3–10 give the results of identification and production for each of the 8 students, respectively. The results are highly consistent across the 8 students. Very little change in performance occurred in both identification and production in the pre-training and post-training sessions, whereas learning occurred during the training sessions.

There was little knowledge or learning of the test items without training, even though these items were repeatedly tested for many days. Once training began on a set of items, performance improved fairly quickly until reaching asymptote. No learning without training and learning with training meet the requirement of the multiple-baseline design. These results support the hypothesis that the training platform was responsible for the learning.

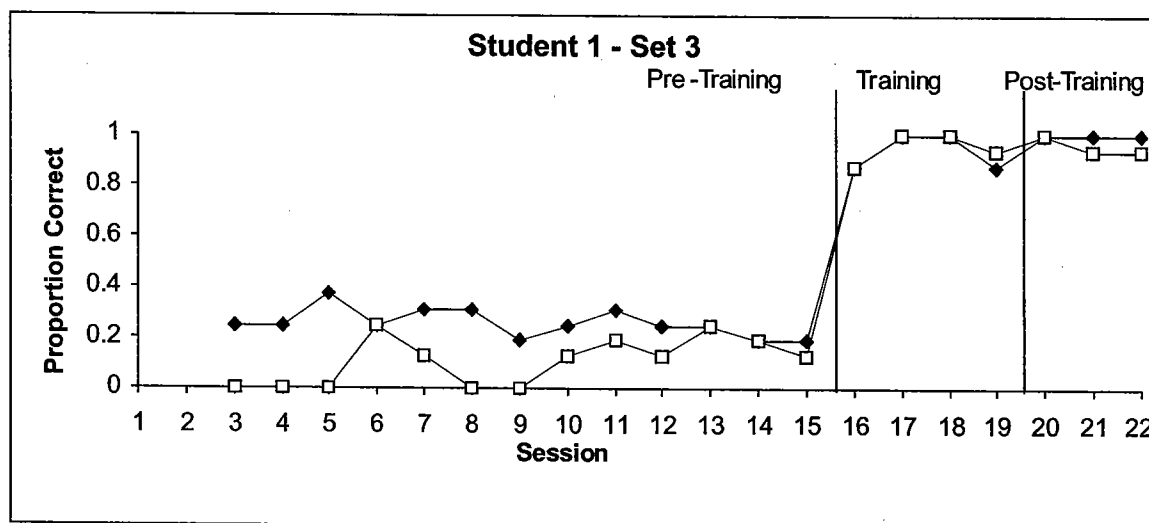
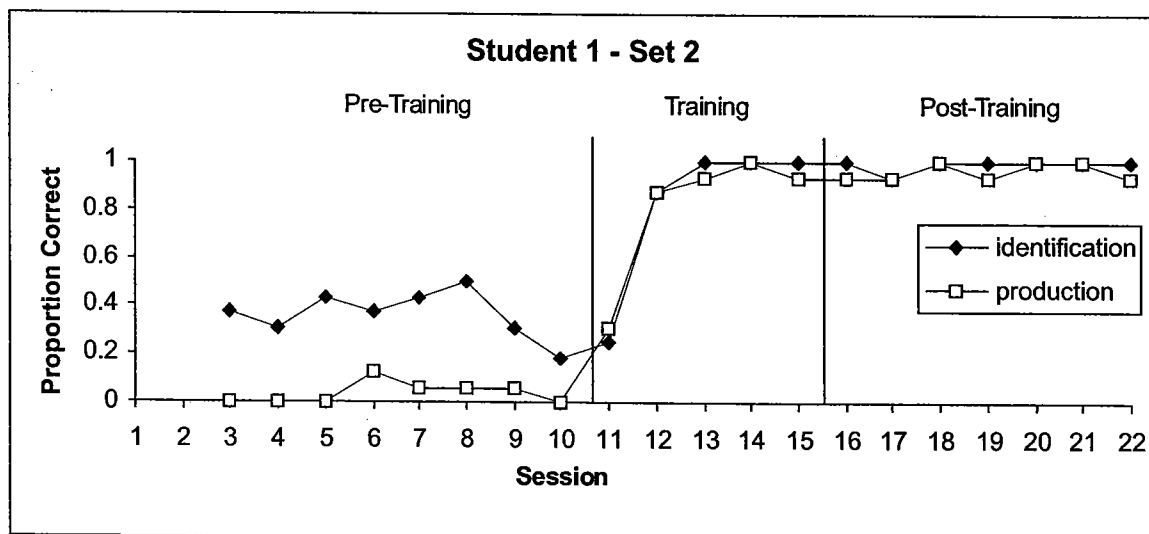
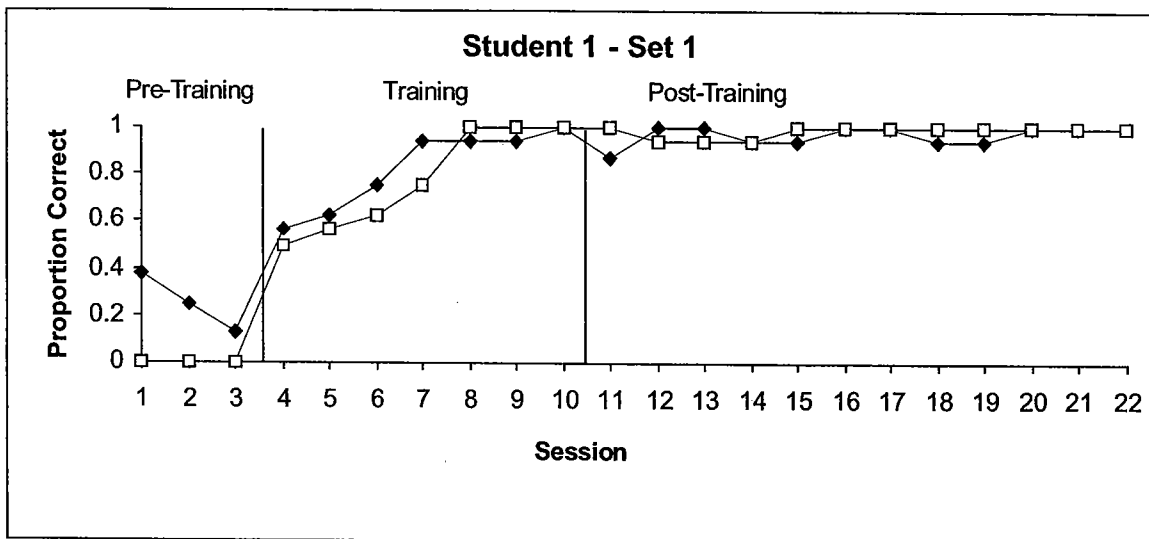


Figure 3. Proportion of correctly identified (black triangles) and correctly produced (white squares) items across the testing sessions, within each set of words, for Participant 1. The training occurred between the two vertical bars. The figure illustrates that very little change occurred in both identification and production in the pre-training and post-training sessions, whereas learning occurred during the training sessions.

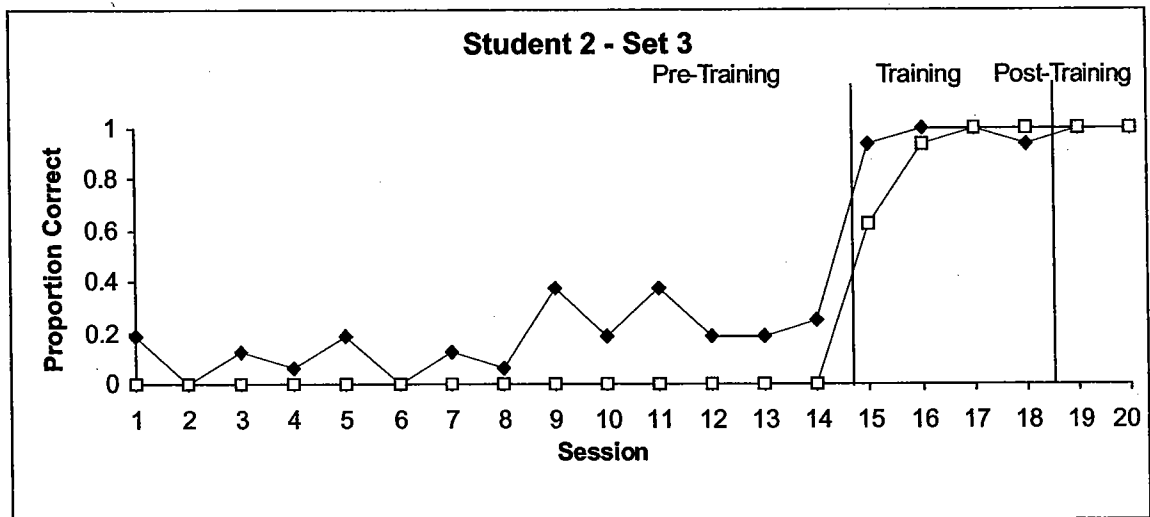
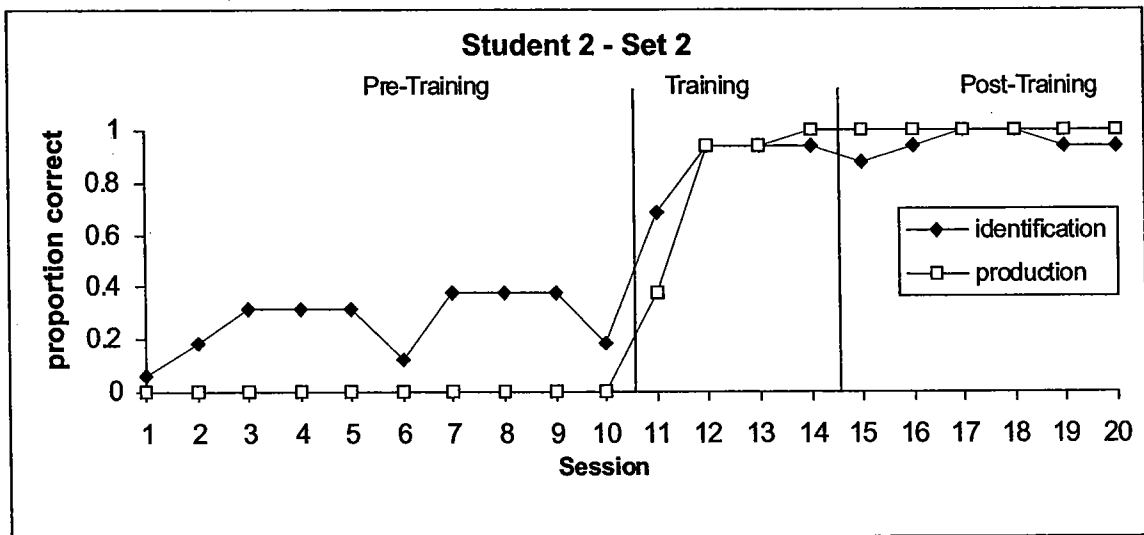
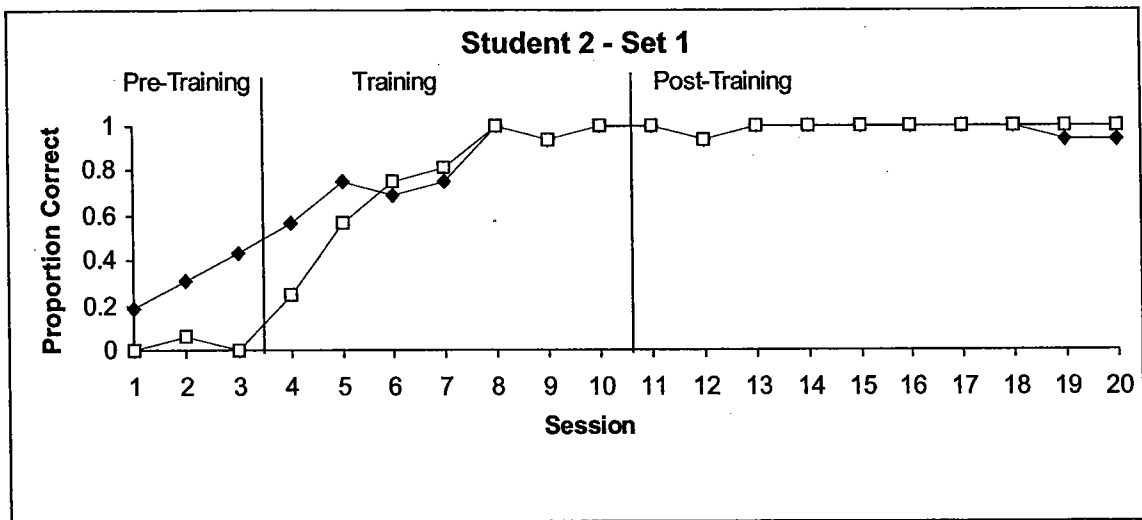


Figure 4. Proportion of correctly identified (black triangles) and correctly produced (white squares) items across the testing sessions, within each set of words, for Participant 2. The training occurred between the two vertical bars. The figure illustrates that very little change occurred in both identification and production in the pre-training and post-training sessions, whereas learning occurred during the training sessions.

Table V. Mean identification and production performance for each participant during pre-training, post-training, and delayed retest.

Participant	Identification- Pre- training	Identification- Post- Training	Identification- Delayed Retest	Production- Pre- training	Production- Post- Training	Production- Delayed Retest
1	.29	.99	.94	.05	.97	.96
2	.25	.98	.98	.01	1.00	1.00
3	.42	.98	.90	.10	.98	.92
4	.30	.99	.98	.07	.99	1.00
5	.12	1.00	.98	.00	1.00	1.00
6	.18	.97	.98	.05	.99	1.00
7	.15	1.00	1.00	.09	1.00	1.00
8	.29	1.00	.98	.05	.99	.98
Mean	.25	.99	.97	.05	.99	.98

Identification accuracy (mean = .72 across all training sessions) was always higher than production accuracy (mean = .64 across all training sessions), $F(1,7) = 25.67$, $p < .001$. This result is not unexpected because a student could associate a name with an image without being able to pronounce it correctly. In addition, guessing accuracy without any knowledge in the identification task is 12.5%, whereas it would be near zero in the production task.

Each student showed a significant improvement after just one day of training. To test whether learning improved significantly from the last day of the pre-training assessment to the assessment given after the first day of training, an analysis of variance was conducted on identification and production accuracy performance on all word sets. The independent variables were identification versus production, word set 1, 2, and 3, and the last day of pre-training versus first day of training. Average accuracy averaged across identification and production improved from .173 on the last day of assessment to .681 after the first day after training, $F(1,7) = 127.54$, $p < .001$. Identification was significantly higher than production $F(1,7) = 9.63$, $p < .05$, and there was no significant interaction between training and type of response.

Figures 3–10 show that performance did not degrade after training on each set of words ended and training on other words took place. In addition, a delayed reassessment test given approximately 4 weeks after completion of the experiment revealed that the students retained the items that were learned. Table 5 gives the proportion of words correct on the pre-training, post-training, and delayed tests for identification and production. As can be seen in the table, all of the students mastered all of the words, and retained these words 4 weeks after the training and post-testing were complete.

Generally, all of the participants showed the same pattern of learning with the first set of words taking the longest to master and each set thereafter being learned at a faster rate. To test the hypothesis that learning rate improved from each set to the next, a two-way analysis of variance was carried out on

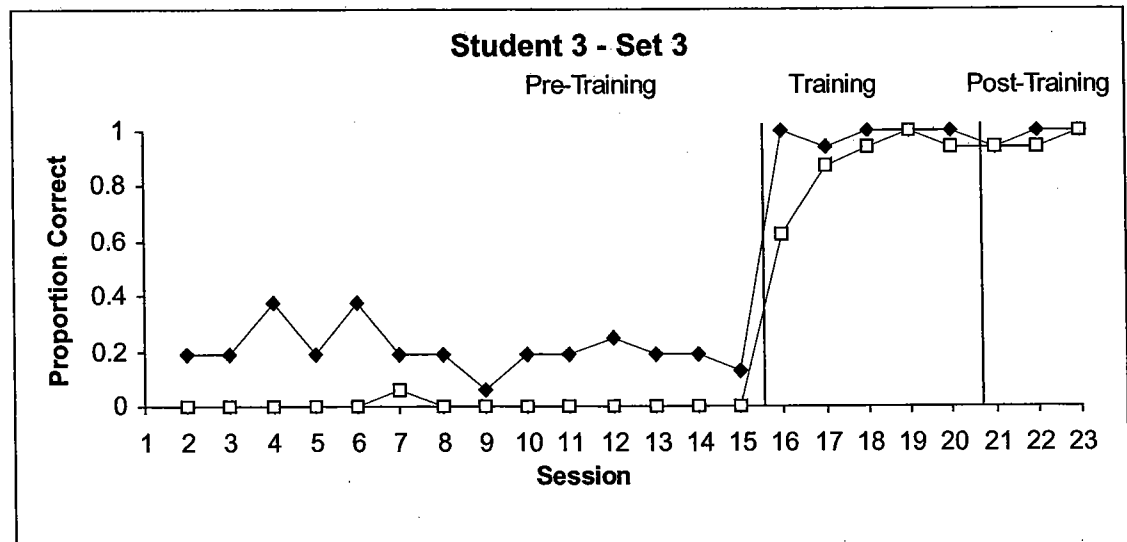
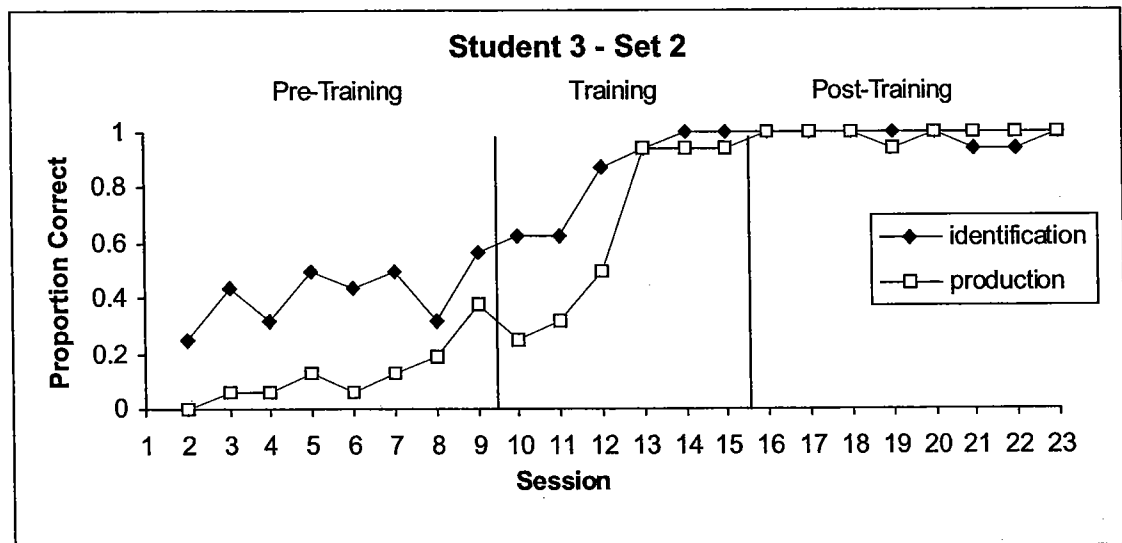
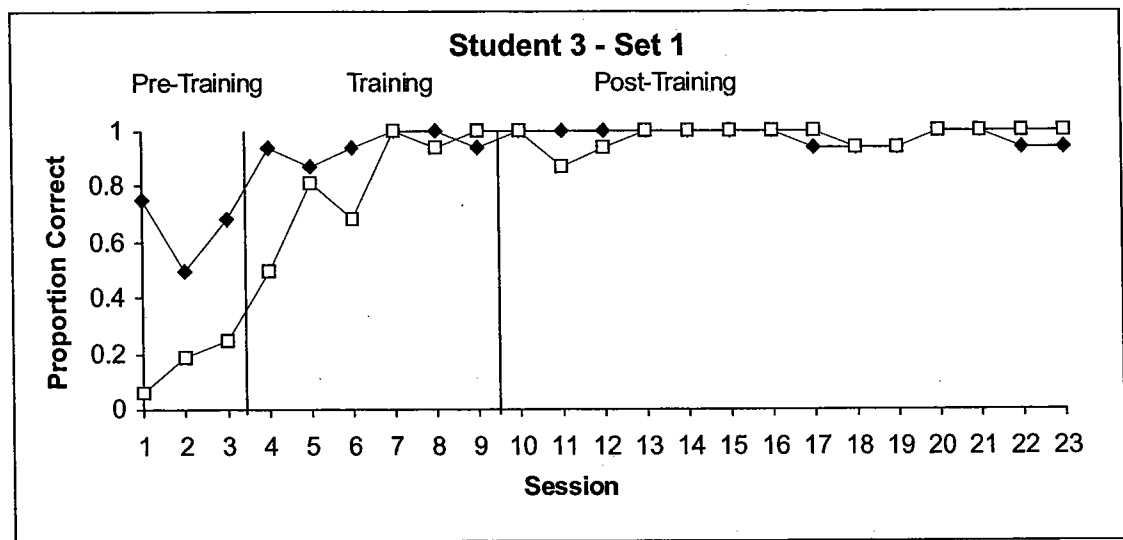


Figure 5. Proportion of correctly identified (black triangles) and correctly produced (white squares) items across the testing sessions, within each set of words, for Participant 3. The training occurred between the two vertical bars. The figure illustrates that very little change occurred in both identification and production in the pre-training and post-training sessions, whereas learning occurred during the training sessions.

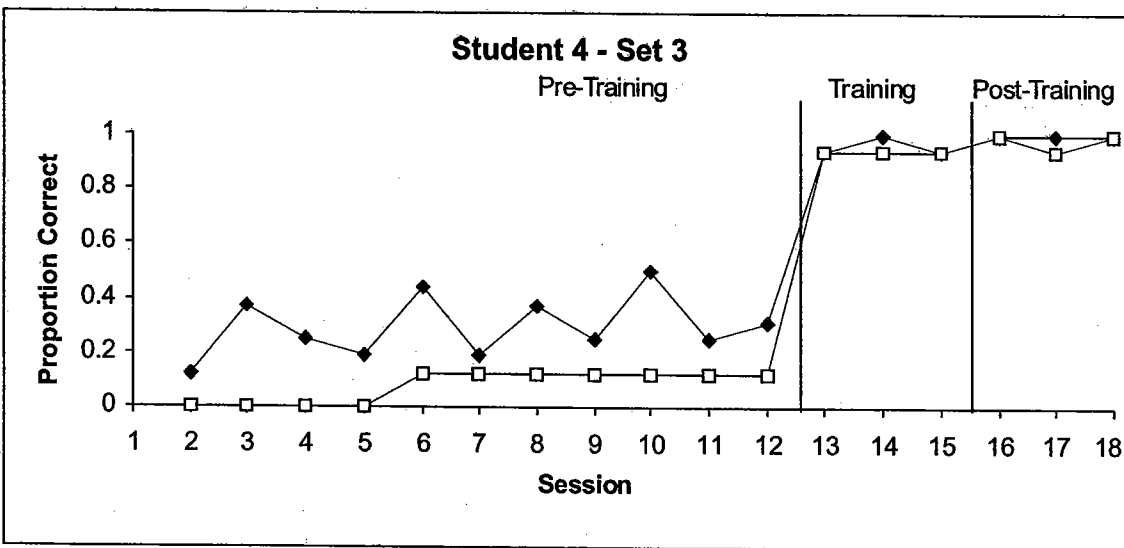
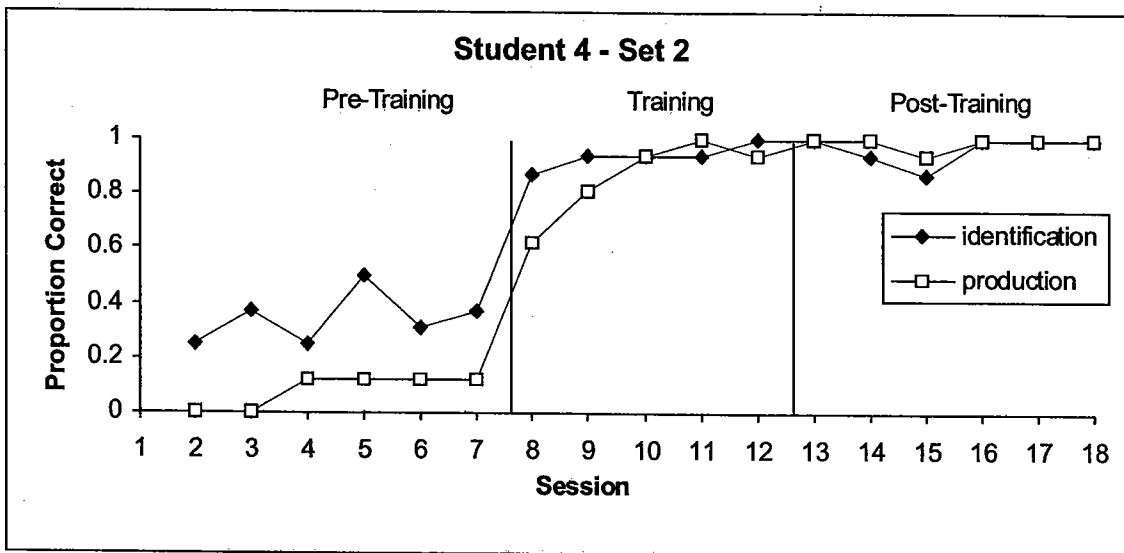
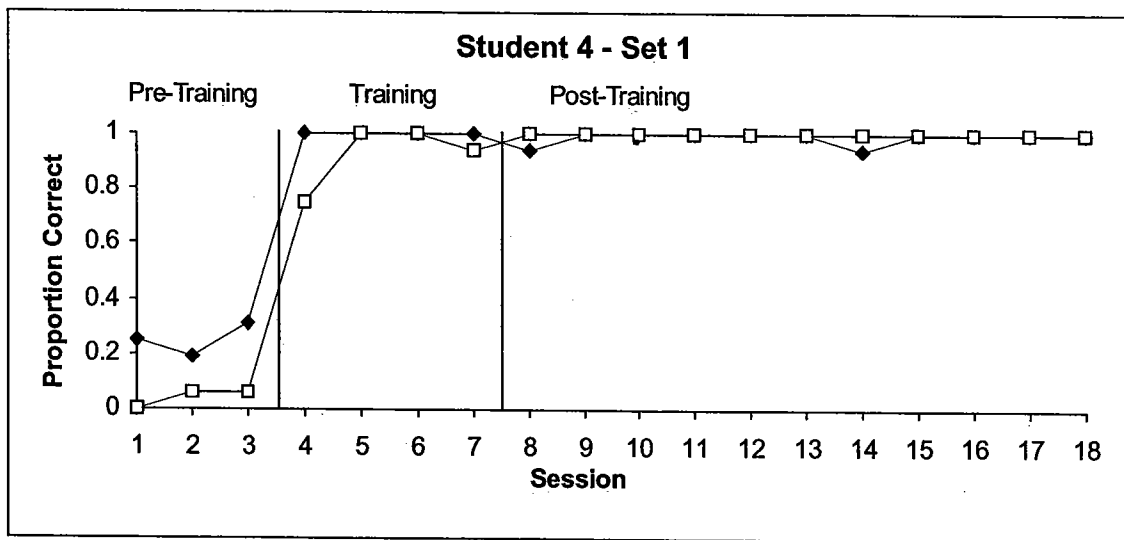


Figure 6. Proportion of correctly identified (black triangles) and correctly produced (white squares) items across the testing sessions, within each set of words, for Participant 4. The training occurred between the two vertical bars. The figure illustrates that very little change occurred in both identification and production in the pre-training and post-training sessions, whereas learning occurred during the training sessions.

the number of sessions that were required to reach a criterion level of at least .94 for identification and production for 3 days in a row to master the first, second, and third sets of words. The analysis of variance showed a significant effect of set, $F(2, 14) = 7.988, p < .05$, and a significant linear trend, $F(1,7) = 14.831, p < .05$. The number of sessions required to reach criterion for identification was significantly higher than production, $F(1,7) = 5.5, p < .05$, but there was no significant interaction between word set and type of response.

Given the scoring criterion for production accuracy, any change in the fine-grained quality of the productions cannot be seen in production accuracies given in Figures 3–10. The experimenter noted, however, that not only accuracy but also the quality of the participants' productions improved over the course of repeated testing and training. In addition to Baldi reinforcing the pronunciation of the words, we believe that the reading exercises were valuable for all of the students as were spelling exercises for the six older students. With repeated exposure to the pronunciation and spelling of the items, the experimenter observed that the participants became more familiar with them and in turn were able to pronounce them more precisely.

The number of trials required to reach criterion was 5, 4.3, and 3.4 for mastering the first, second, and third sets of categories. Although many categories involved in training overlapped from one student to the next, the items that made up these categories were unique to each participant (see Table 2). Given that the word lists were randomized across participants, the differences in the difficulty of the word sets are probably not responsible for this apparent learning-to-learn process. An obvious explanation for this facilitation is that the children would have had more testing sessions for the second and third sets of words. These testing sessions would allow the children to learn the names of the words without any association with the image. Thus, the nature of the testing and training procedure may have contributed this apparent learning-to-learn process. The participants may have also become increasingly more comfortable and familiar with Baldi and the method of training, which could have facilitated the learning of the second and third sets.

When testing began, the students seemed uneasy about not knowing the items being presented. They either didn't respond at all or they responded with "I don't know" to Baldi's prompts of "what do you call this?" or "what is this?" and this elicited frustration. A parent of one student sat in on an early training session and was astonished to learn that her child did not know very basic everyday labels for common objects such as certain fruits and vegetables. The students' teachers also expressed the same concern. Parents often incorrectly assume that their child knows the labels for common household items. Many instances during testing and training proved that the student had seen the item before, for they would say, "I forget what this is called" or "I have seen that before," but they weren't able to identify the image or produce the word.

As training on the first set of words began, the students quickly realized that they could easily learn these items and enjoyed tracking their progress through the feedback provided during training. The students could individ-

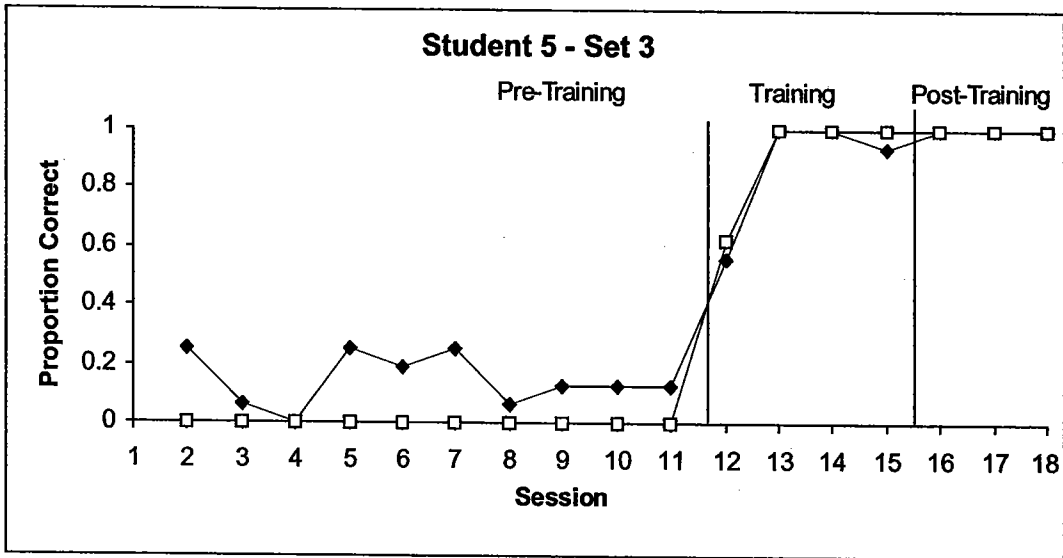
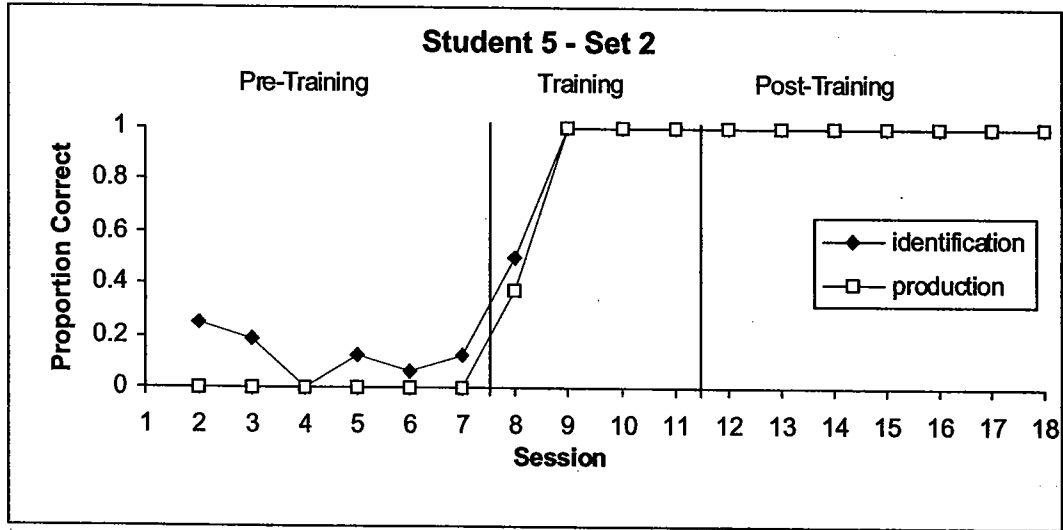
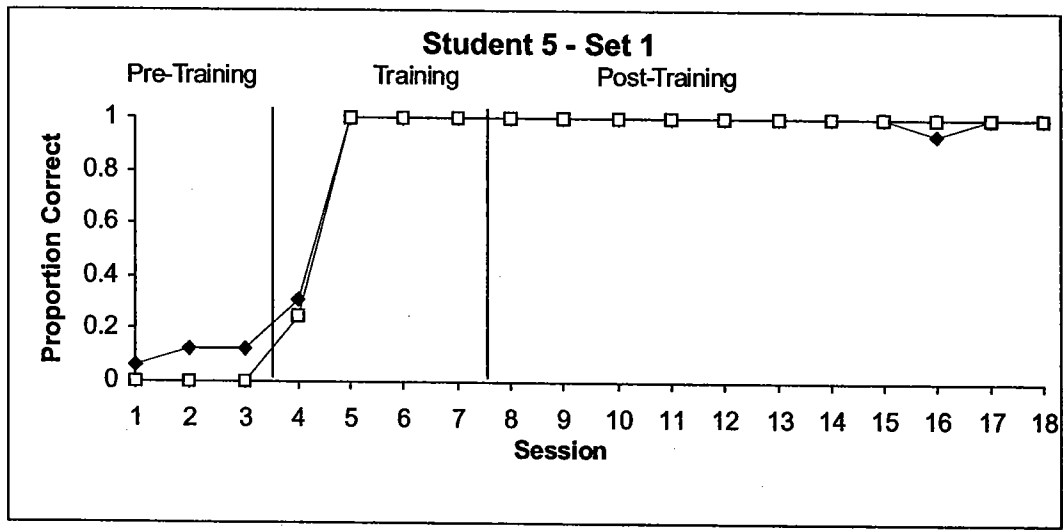


Figure 7. Proportion of correctly identified (black triangles) and correctly produced (white squares) items across the testing sessions, within each set of words, for Participant 5. The training occurred between the two vertical bars. The figure illustrates that very little change occurred in both identification and production in the pre-training and post-training sessions, whereas learning occurred during the training sessions.

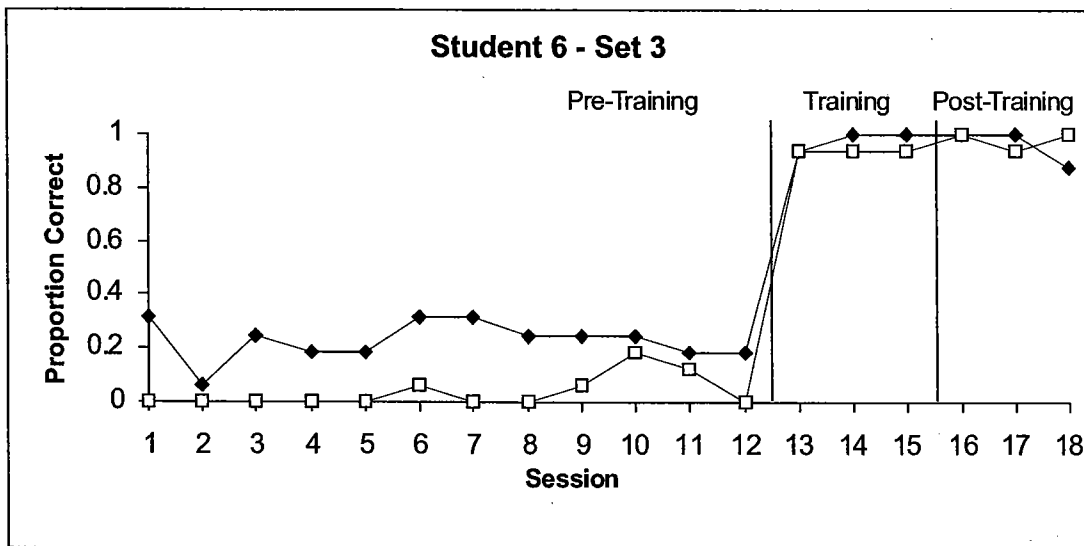
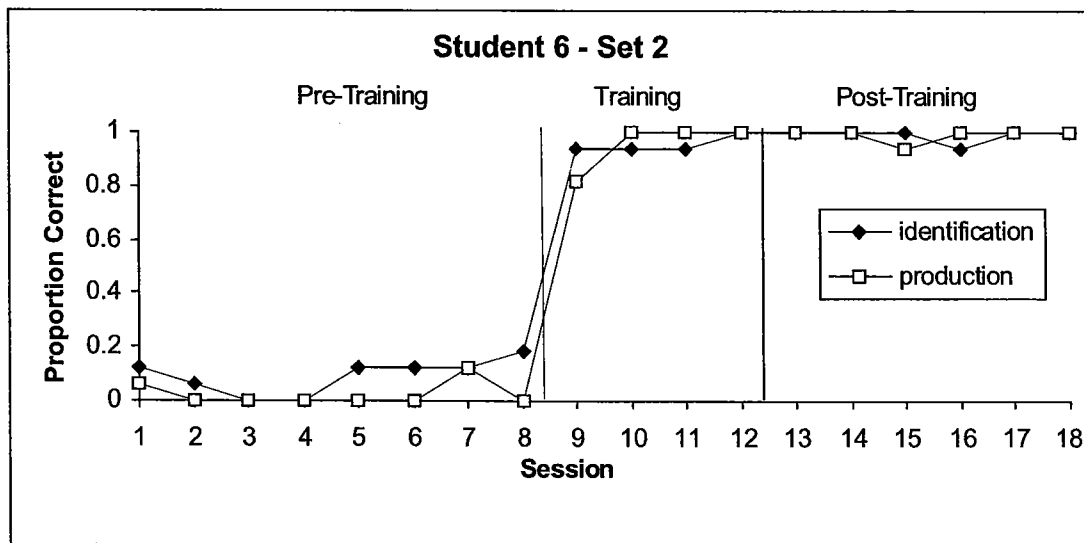
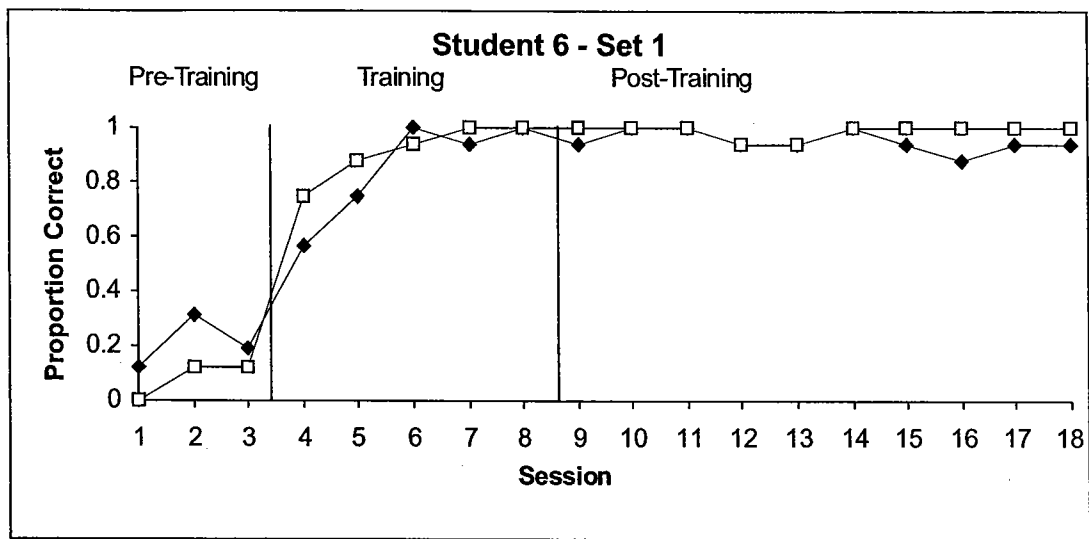


Figure 8. Proportion of correctly identified (black triangles) and correctly produced (white squares) items across the testing sessions for, within each set of words for Participant 6. The training occurred between the two vertical bars. The figure illustrates that very little change occurred in both identification and production in the pre-training and post-training sessions, whereas learning occurred during the training sessions.

ually note when they had mastered a set of items through the feedback and were eager to move on to the next set of words. Many of the students wanted to know when they would be able to start the next set of training items. They always wanted to learn more.

When a set of training items was completed, many students found themselves bored with the testing of that mastered set. They were more interested in learning words they did not know rather than repeating items they already felt they knew. The children also became somewhat impatient when feedback was withheld on the assessment trials. When the third and final set of words was being trained, the students seemed to be more confident in their responses, for they no longer had to be embarrassed by responding with "I don't know."

Given that the design of this experiment was a within-participant control, individual performance of the eight students (with their ages and auditory device thresholds in Table 1) merits consideration. Participants fell into one of two age groups: Participants 1 and 2 were 6–7 years old and Participants 3–8 were 9–10 years old. Participant 1 was one of the younger participants in the study. In addition, his teacher expressed that this student needed a lot of repetition to learn. Thus, it took him a bit longer to learn the words than most of the other students. He required about seven training sessions to learn the first set of words, about five sessions to learn the second set, and about four sessions to learn the third. Although it took this student many sessions to reach the criterion level for training on the first set, the number of training sessions involved in the subsequent word sets was less. Participant 2 was also younger than the rest of the participants. Like Participant 1, learning occurred more quickly for the second and third sets of words. This student was able to learn the labels to the categories on a whole very quickly, but had particular trouble remembering certain items, for example, he would confuse cabbage and raddish. This student had a congested nose during many of the sessions, so his congestion may have contributed to his hearing difficulty as well.

Participant 3 was very fond of this program when it started, but quickly became bored with the repetition of testing and training sessions. She started to lose focus and motivation as time went on, primarily because of the large amount of testing with no training on two-thirds of the words at any one time. No increased rate of learning for this student might be explained by a decrease in concentration. In order to retain this student's motivation, the experimenter promised her a reward upon finishing the study with all words correct. This seemed to increase her attention span somewhat.

Participant 4 was bilingual in English and Spanish and was always very attentive and eager to learn. She enjoyed working with Baldi and continuously asked the experimenter when she was going to be able to learn the unknown items. For the Set 2 words, she had a hard time distinguishing the difference between two pairs of animals (hyena vs. iguana, antelope vs. anteater) and pronouncing these words as well. This might explain why Set 2 was not learned as quickly as the others.

ed
or
es
ig

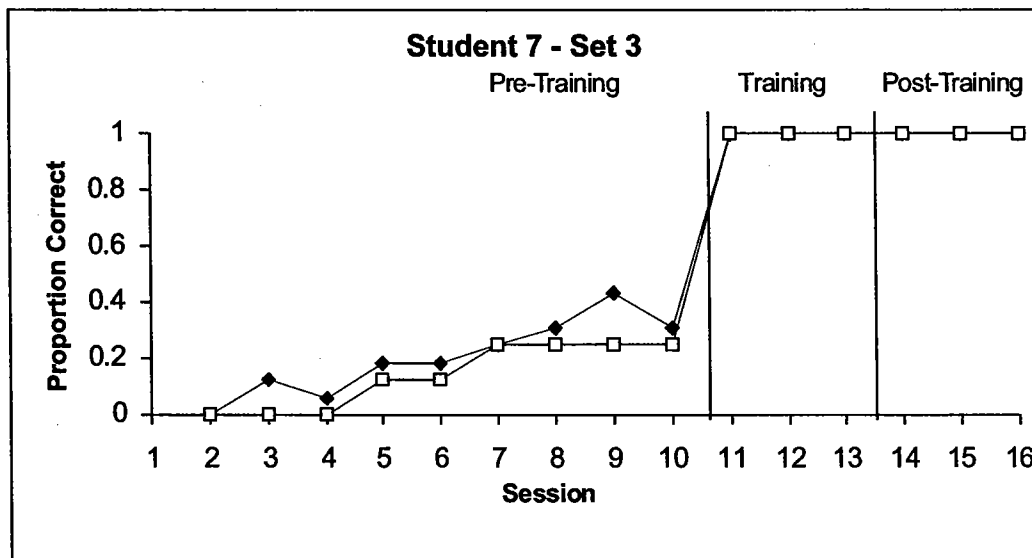
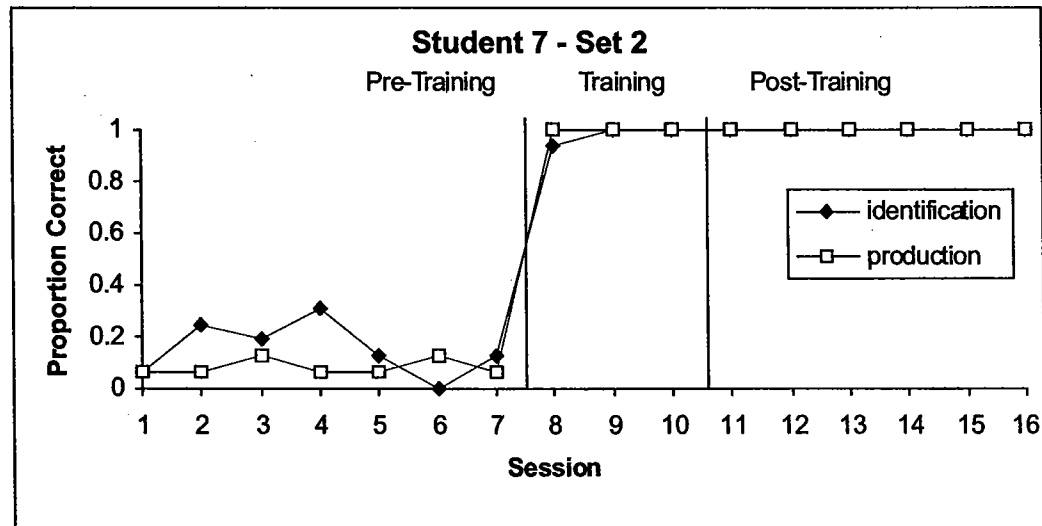
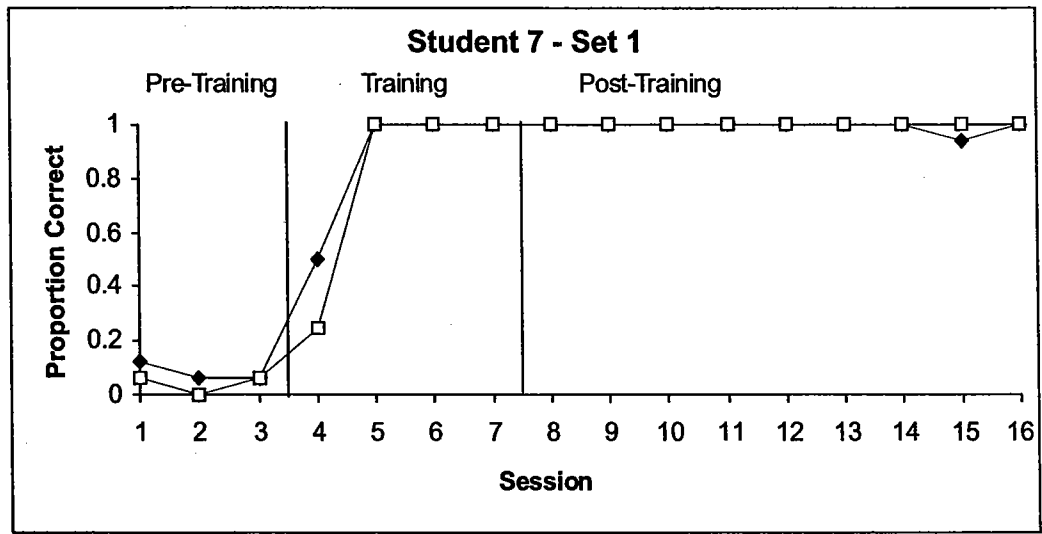


Figure 9. Proportion of correctly identified (black triangles) and correctly produced (white squares) items across the testing sessions, within each set of words for Participant 7. The training occurred between the two vertical bars. The figure illustrates that very little change occurred in both identification and production in the pre-training and post-training sessions, whereas learning occurred during the training sessions.

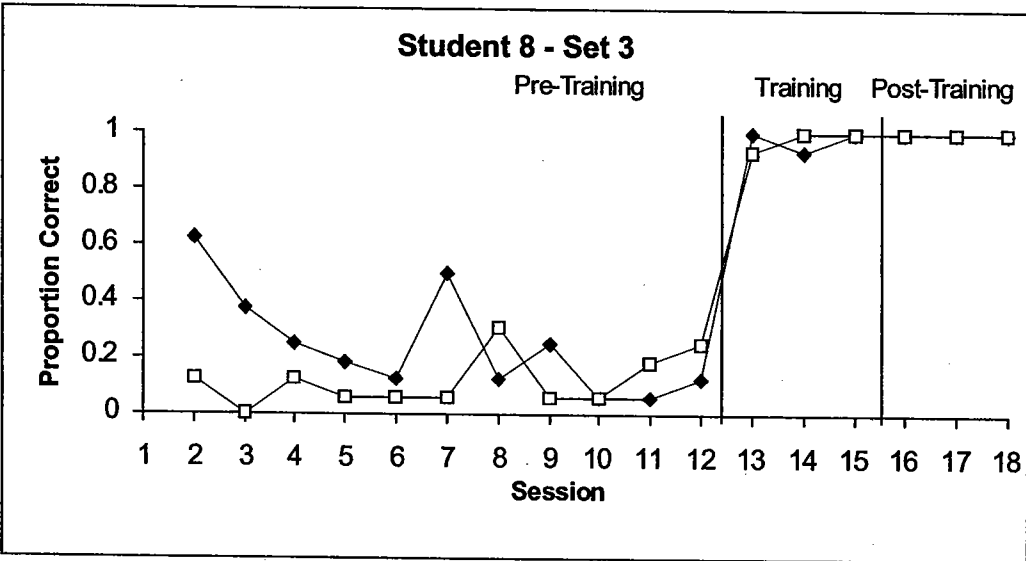
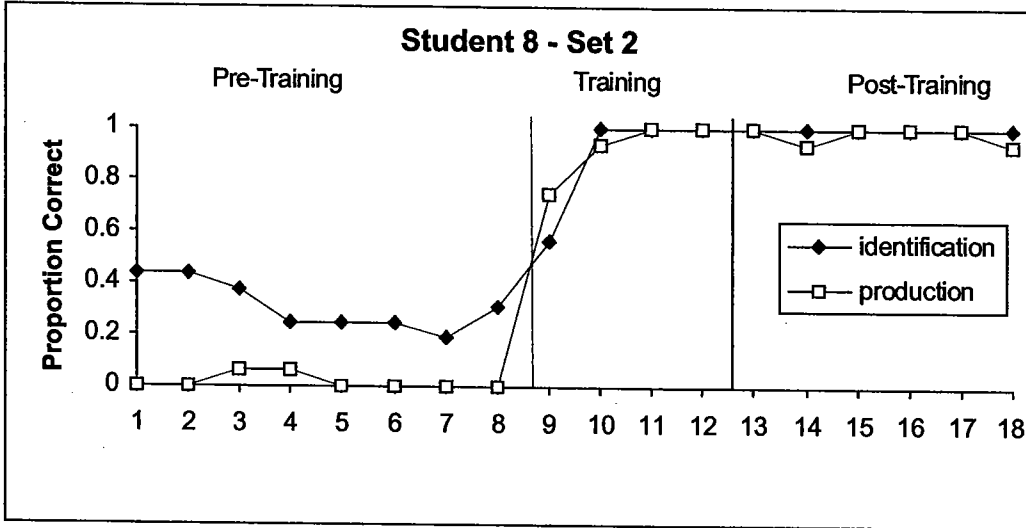
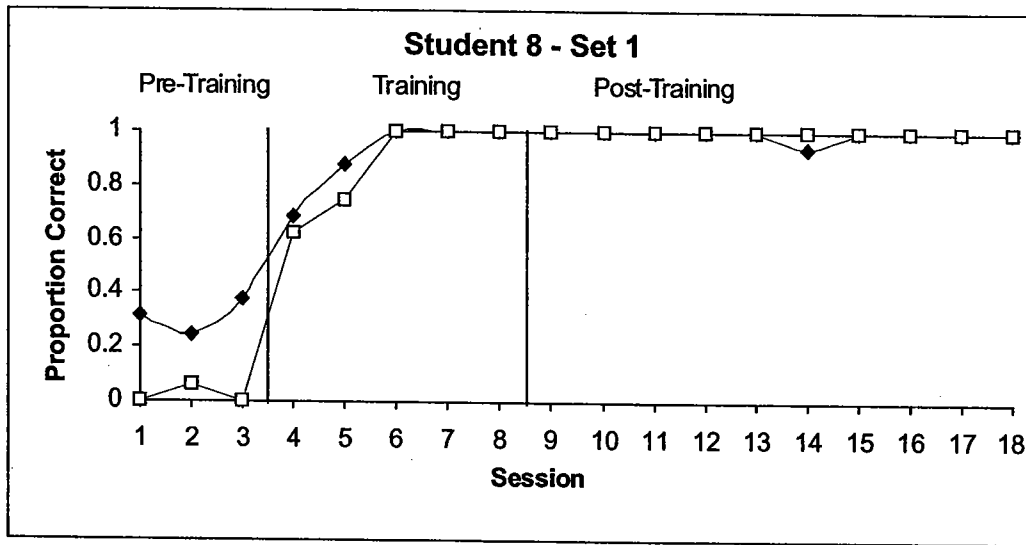


Figure 10. Proportion of correctly identified (black triangles) and correctly produced (white squares) items across the testing sessions, within each set of words for Participant 8. The training occurred between the two vertical bars. The figure illustrates that very little change occurred in both identification and production in the pre-training and post-training sessions, whereas learning occurred during the training sessions.

