Published in final edited form as:

J Speech Lang Hear Res. 2008 February ; 51(1): 97–119.

The Efficacy of Fast ForWord-Language Intervention in School-Age Children with Language Impairment: A Randomized Controlled Trial

Ronald B. Gillam,
Utah State University

Diane Frome Loeb,
The University of Kansas

LaVae M. Hoffman,
University of Oklahoma Health Sciences Center

Thomas Bohman,
The University of Texas at Austin

Craig A. Champlin,
The University of Texas at Austin

Linda Thibodeau,
The University of Texas at Dallas

Judith Widen,
The University of Kansas

Jayne Brandel, and
The University of Kansas

Sandy Friel-Patti
The University of Texas at Dallas

Abstract

Purpose—A randomized controlled trial (RCT) was conducted to compare the language and auditory processing outcomes of children assigned to Fast ForWord-Language (FFW-L) to the outcomes of children assigned to nonspecific or specific language intervention comparison treatments that did not contain modified speech.

Method—Two hundred and sixteen children between the ages of 6 and 9 years with language impairments were randomly assigned to one of four arms: Fast ForWord-Language (FFW-L), academic enrichment (AE), computer-assisted language intervention (CALI), or individualized language intervention (ILI) provided by a speech-language pathologist. All children received 1 hour and 40 minutes of treatment, 5 days per week, for 6 weeks. Language and auditory processing measures were administered to the children by blinded examiners before treatment, immediately after treatment, 3 months after treatment, and 6 months after treatment.

Results—The children in all four arms improved significantly on a global language test and a test of backward masking. Children with poor backward masking scores who were randomized to the FFW-L arm did not present greater improvement on the language measures than children with poor backward masking scores who were randomized to the other three arms. Effect sizes, analyses of standard error of measurement, and normalization percentages supported the clinical significance of

the improvements on the CASL. There was a treatment effect for the Blending Words subtest on the *Comprehensive Test of Phonological Processing* (Wagner, Torgesen, & Rashotte, 1999). Participants in the FFW-L and CALI arms earned higher phonological awareness scores than children in the ILI and AE arms at the six-month follow-up testing.

Conclusion—Fast ForWord-Language, the language intervention that provided modified speech to address a hypothesized underlying auditory processing deficit, was not more effective at improving general language skills or temporal processing skills than a nonspecific comparison treatment (AE) or specific language intervention comparison treatments (CALI and ILI) that did not contain modified speech stimuli. These findings call into question the temporal processing hypothesis of language impairment and the hypothesized benefits of using acoustically modified speech to improve language skills. The finding that children in the three treatment arms and the active comparison arm made clinically relevant gains on measures of language and temporal auditory processing informs our understanding of the variety of intervention activities that can facilitate development.

Approximately seven percent of all school-age children have unusual difficulty learning and using language despite adequate hearing, nonverbal intelligence, and motor abilities (Tomblin, Records, & Zhang, 1996). This difficulty, which has been referred to by a variety of terms including language impairment, language-learning disability, specific language impairment, and language-learning impairment, can have serious social, academic, and vocational ramifications (Brinton, Spackman, Fujiki, & Ricks, 2007; Catts, Fey, Tomblin, & Zhang, 2002; Clegg, Hollis, Mawhood, & Rutter, 2005; Conti-Ramsden & Botting, 2004). Unfortunately, a high percentage of school-age children with language impairments do not make gains beyond one standard error of measurement on formal tests administered one, three, or even as many as 10 years after they were identified (Aram & Hall, 1989; Bernstein & Stark, 1985; Conti-Ramsden & Botting, 1999; Fazio et al., 1996; Silva et al., 1987; Snowling & Hulme, 1989; Tallal et al., 1996; Tomblin, Zhang, Buckwalter, & O'Brien, 2003). While only indirectly a measure of treatment efficacy, such results are not a strong endorsement of the language intervention services that have been provided to school age children.

One popular approach to language intervention for school-age children is Fast ForWord-Language™ (FFW-L, Scientific Learning Corporation, 1998). FFW-L has received a great deal of attention in the scientific literature and the lay press. The developers of FFW-L assert that the program leads to neural reorganization that causes an increased ability to perceive fast changing acoustic input, and that such improvement in leads to subsequent gains of one to one-and-one-half years on standardized tests of language skills after six weeks of training (Merzenich et al., 1996; Tallal, 2004; Tallal and Gaab, 2006; Tallal et al., 1996).

Two hypotheses underlying the use of FFW-L in intervention are that language impairments (LI) result from difficulty recognizing and sequencing the spectrotemporal structure of speech (Tallal, 1980; 1990; 2004) and that temporal processing abilities improve as a result of computer-assisted instruction in which modifications to the acoustic and temporal properties of the speech signal are gradually reduced as a function of increased performance on sound, syllable, word, and sentence comprehension tasks (Agocs, Burns, de Ley, Miller, & Calhoun, 2006; Merzenich et al., 1996; Tallal et al., 1996). Merzenich, Tallal, and their colleagues published two influential studies about changes in temporal processing and language comprehension in children with language impairments (Merzenich et al., 1996; Tallal et al., 1996). In the first study, 7 children with LLD between 5 and 9 years of age received language intervention 3 hours per day, 5 days per week, for 4 weeks. The participants were characterized as having normal nonverbal intellectual abilities, delays in receptive and expressive language development, and reading difficulties. The children played prototypes of the FFW-L games called Circus Sequence and Phoneme Identification, but they also rotated through eight other speech and language exercises that were presented by trained clinicians in individual sessions. In addition, children completed 1 to 2 hours of homework every day that involved listening to

stories. After training, the children improved significantly on measures of speech discrimination and language comprehension.

A second study (Merzenich, et al., 1996; Tallal, et al., 1996) was conducted with 22 children with language impairments who ranged in age from 5;4 to 10;0 years. These children were divided into two matched groups according to nonverbal intelligence and receptive language abilities. Children in both groups attended laboratory sessions for 3.5 hours each day. The children in the second study also completed 1 to 2 hours of listening homework each day. The groups differed according to the auditory stimuli that were presented. Children in one group listened to modified speech as they played revised versions of the Circus Sequence and Phoneme Identification games that were used in Experiment 1 and two additional games (Old MacDonald's Flying Farm and Phonic Match) that are part of the current FFW-L package. They also listened to modified speech in the clinician-directed intervention sessions each day and in their daily homework sessions. The children in the second group, "...received equivalent language training but with natural speech materials," and they "...played video games rather than these adaptive auditory-speech training games" (Merzenich et al., 1996, p. 80). After training, children in both treatment groups improved significantly on all measures. The children whose language intervention included modified speech stimuli evidenced greater improvement on measures of temporal processing, speech discrimination, and grammatical comprehension than children in the natural speech treatment group. The authors concluded that the training remediated the underlying temporal processing deficit that contributed to difficulties with speech perception and language comprehension (Merzenich et al., 1996; Tallal et al., 1996).

Despite decades of providing language intervention to school age children, there are few studies comparing outcomes of different intervention procedures. The absence of such data is perhaps one of the reasons for the public and professional excitement about the studies of FFW-L reported by Tallal and colleagues (Merzenich, et al., 1996; Tallal, et al., 1996). Parents and professionals have shown keen interest in data demonstrating that FFW-L intervention is associated with dramatic improvement in language test scores.

Two independent groups of researchers have evaluated the efficacy of FFW-L with children who have language impairments. Pokorni, Worthington, and Jamison (2004) randomly assigned a group of fifty-four children with language and reading impairments to one of three interventions: FFW-L, Earobics, or the Lindamood Phoneme Sequencing Program (LiPS). The children scored at least one standard deviation below the mean on one of three subtests of the Clinical Evaluation of Language Functions-3 (CELF-3; Semel, Wiig & Secord, 1995): Concepts and Directions, Recalling Sentences, or Listening to Paragraphs, and had reading skills that were at least one year below their grade level. The children were primarily African American (75%) and Caucasian (25%), and approximately 42% were living in families that were at or below the level of poverty. Children in all three groups received three, 1-hour intervention sessions each day for 20 days. Testing occurred 4–6 weeks prior and 6–8 weeks after the completion of intervention. Pokorni et al. (2004) found no significant gains on the three subtests of the CELF– 3 for any of the treatments.

Cohen, et al. (2005) studied the effectiveness of FFW-L in a randomized controlled trial (RCT) that included seventy-seven children with severe mixed receptive and expressive language impairment. On average, the children scored at least 2 standard deviations below the mean on both the receptive and expressive subtests of the Clinical Evaluation of Language Fundamentals, CELF-3 UK version (Semel, Wiig & Secord, 2000). Children were randomly assigned to one of three groups: one that received home intervention with FFW-L, another that received home intervention with commercially available language and reading computer games that did not contain modified speech, or a third group that received no home intervention. All the children received regular speech-language intervention services at school during the

study. There were between 23–27 children in each group. The children in all three groups made significant gains on the CELF-3 UK at 9 weeks and at 6 months after intervention. However, there were no significant differences among the groups. Thus, it was concluded that the FFW-L intervention did not provide any additional benefit for children with severe language impairments above the current language intervention services they were receiving at school. Thus, their findings called into question the utility of using acoustically modified speech to remediate severe, mixed language impairments.

There are some important limitations to the previous clinical trials that tested the language outcomes of FFW-L. Both trials included a relatively small subset of the children who typically receive treatment for language impairments. Cohen et al. studied only those children with mixed receptive and expressive language impairments; Pokorni et al. studied children with language and reading disorders. In addition, none of the previous clinical trials measured changes in temporal auditory processing directly.

Our study was designed to overcome some of the limitations of the previous investigations. The study was undertaken over the period of three years and included a total of 216 children, almost three times the number of children in the Cohen et al. study and four times the number of children in the Pokorni et al. study. In our RCT, we compared FFW-L to two other computer instruction conditions: a computer assisted language intervention (CALI) condition in which children played language intervention programs that specifically targeted cognitive, processing, and language skills similar to the ones targeted by FFW-L but without a modified speech signal, and an academic enrichment (AE) condition in which children played education computer games that were not specifically designed to improve language skills. The academic enrichment condition was designed to be an active comparison. That is, the AE condition shared a number of important features with the FFW-L and CALI conditions. However, the computer games in the AE condition were not specifically designed to improve language skills or auditory processing skills. We also included an individual language intervention (ILI) condition that was delivered by a speech-language pathologist. Because we held frequency and duration of intervention constant, we were able to compare three critical dimensions of intervention: (a) computer-delivered versus human-delivered services; (b) modified speech versus unmodified speech; and (c) specific versus nonspecific intervention goals.

The primary research questions were:

1. Will participants randomly assigned to the FFW-L arm show greater improvement on the CASL composite score than participants assigned to either of two specific intervention arms (CALI or ILI) or to a nonspecific comparison (AE) arm immediately after intervention and on follow-up testing three and six months later?
2. Will participants randomly assigned to the FFW-L arm show greater improvement on a measure of backward masking than the participants assigned to the specific intervention arms (CALI or ILI) or to the nonspecific comparison (AE) arm immediately after intervention and on follow-up testing three and six months later?

We reasoned that if the auditory temporal processing deficit theory of language impairment is correct, then participants who received the FFW-L intervention should show more improvement on language and auditory processing measures than the children who were randomized to the other treatments or to an active comparison group that played academic enrichment computer games. Following this logic, the FFW-L advantage should be maintained or strengthened three months and six months after treatment. The design of our study also enabled us to explore two secondary questions: Is the modified speech in the FFW-L intervention necessary for language improvement to occur? Further, are computer-delivered services that do not require a verbal response as efficacious as human-delivered services that include verbal interaction?

Method

Participants

Recruitment and Identification—Two hundred sixteen children with language impairments (LI) participated in the study at three different sites. The children were recruited from 9 school districts, with 96 children from Northeast Kansas, 92 children from Central Texas, and 28 children from North Texas. Between January and May of each year, children were recruited and tested to determine if they qualified for inclusion in the study. The investigators met with speech-language pathologists, teachers, and special educators to explain the purpose of the investigation and the inclusion criteria. School district personnel gave brochures about the study to the parents of children whom the educators thought might meet the eligibility criteria. Parents contacted the research coordinators in each area if they were interested in having their child participate. After a brief conversation with the parents, the research coordinator scheduled an independent assessment for those children who appeared to be potential participants.

Consistent with the EpiSLI model (Tomblin et al., 1997), children were determined to have a language impairment if they displayed a standard subtest score between 75 and 125 (± 1.66 SD) on the Matrices subtest of the Kaufman Brief Intelligence Test (K-BIT) (Kaufman & Kaufman, 1990) together with standard scores at or below 81 on two or more clusters of the Test of Language Development: Primary: 3rd edition (TOLD:P:3) (Newcomer & Hammill, 1997). To be eligible, participants could not present a hearing impairment, a visual impairment, gross neurological impairment, oral-structural anomalies, or emotional/social disorders. Vision and hearing status was confirmed either through school records of passing screenings within the past year or through live administration of vision and/or hearing screenings. When school vision screening records were unavailable, a vision screening was completed using Lea Symbols vision screening materials to ensure adequate vision in at least one eye with or without corrective lenses. Hearing screenings were administered at 20 dBHL at the frequencies of 1, 2, and 4 KHz in both ears. Oral structure and function screening with the Oral Speech Mechanism Screening Exam-3 (St. Louis & Ruscello, 2000) ruled out the presence of oral-structural anomalies that could have interfered with normal language production.

Children were excluded if parent responses to questionnaires indicated three or more episodes of otitis media in the previous 12-month period, a history of focal brain lesions, traumatic brain injury, cerebral palsy, seizure disorders, symptoms of severely impaired reciprocal social interaction, or severely restricted activities listed in the DSM-IV criteria for autism spectrum disorders. We also excluded children who had participated in 8 or more hours of language intervention or classroom activities using any of the Fast ForWord, Laureate, or Earobics speech-language or reading software, and/or the Lindamood-Bell auditory discrimination training, because this training may have confounded the results from the current study. In addition, parents agreed not to enroll their child in any other language intervention during the treatment phase of the study.

Randomization—Following identification testing and pretesting, participants were randomly assigned to one of the four arms of the study by the Biostatistics Center at the University of Iowa. Randomization of participants was stratified by treatment site (Austin, Dallas, Kansas City) and SES (high, low). Randomization can minimize type I error and improve statistical power for smaller RCTs (<400 patients) (Kernan et al., 1999). In this study, stratified randomization helped by ensuring that the potential impact of otherwise uncontrollable factors associated with SES were equally distributed across the treatment arms at each site. The treatment assignments were sent to the research coordinators at each site via overnight express mail. The research coordinators were the only individuals who had a record of subject names, subject identification codes, and treatment assignments.

Participant Characteristics—Table 1 shows the pre-intervention means and standard deviations for all participants combined and for children in the four arms. There were 54 children in each arm. Across the four arms, there were more males (136) than females (80); a ratio of 1.7:1. The average age of the 216 participants was 7;6 (years; months) with a range of 6;0 to 8;11. There were 58 six-year-olds, 78 seven-year-olds, and 80 eight-year-olds. The mean nonverbal intelligence score was within normal limits for each treatment group. Prior to intervention, group mean standard scores on the Spoken Language Composite of the *Test of Language Development: Primary – 3rd Ed* (TOLD:P-3; Newcomer & Hammill, 1997) ranged between 69.8 and 73.7, with an overall mean of 73.7. Preliminary ANOVA indicated that none of the group differences reached significance on the identification measures. Overall, the children in the study fit the profile of having normal nonverbal intellectual abilities and language abilities that were consistently 1.2 standard deviations or more below the mean.

There were no differences in the gender distribution across the four arms, $X^2(6, N = 216) = 6.13, p = .41$. Forty-six percent of the children who were enrolled in the study were White-not Hispanic. The next largest race/ethnicity group was Black/African American (29%) followed by children who were White-Hispanic/Latino (15%). The Other category (10%) included children who were Native American, Asian American, more than one race, and children whose parents elected not to report their race or ethnicity. There were no differences in the distribution of race/ethnicity groups across the four arms, $X^2(9, N = 216) = 7.75, p = .56$.

Parent education level was determined based on self-report. There were three possible categories: (1) neither parent had attended college, (2) at least one parent had attended college (including community college), and (3) one or both parents had earned a college degree. In each treatment group, the majority of the children were from families in which at least one parent had attended a university, college, or community college. There was no significant difference between the number of children in the three parental groups across the four arms of the study, $X^2(6, N = 215) = 2.45, p = .87$.

Arms of the Study

Fast ForWord—Language (FFW-L)—Children who were randomly assigned to the FFW-L (Scientific Learning Corporation, 1998) arm played seven different computer games that targeted discrimination of tones (viz., Circus Sequence), detection of individual phoneme changes (viz., Old McDonald's Flying Farm), matching phonemes to a target (viz., Phoneme Identification), identifying matched syllable pairs (viz., Phonic Match), discriminating between minimal pair words (viz., Phonic Words), recalling commands (viz., Block Commander), and comprehending grammatical morphemes and complex sentence structures (viz., Language Comprehension Builder).

The speech and nonspeech stimuli in the FFW-L computer games were modified by an algorithm that prolonged segments and differentially amplified particular frequencies (Nagarajan et al., 1998). The acoustic modifications were gradually decreased as children improved on each task. Children received trial-by-trial feedback for correct and incorrect responses, and the movement through the program was controlled automatically. When children responded incorrectly, the correct answer was provided prior to the next stimulus. Correct responses were rewarded by points, jingles, and extra animations on the computer screen.

The children played the games in a quiet area and wore Sony MDR-V-150 closed, supra-aural dynamic stereo headphones, which were used for all three computer arms. A certified and licensed speech-language pathologist supervised every session and graduate student monitors were present in the room with the children and were encouraged to provide the child with positive nonverbal signals (i.e., thumbs up, smiles) to encourage continued work. Five of the

seven games were presented each day for 20-minute periods (a total of 1 hour and 40 minutes per day) until the child reached a criterion of 90% completion on any five games. Even though a few children reached this criterion, they continued playing the games at the highest level through the last day of the intervention session.

Computer Assisted Language Intervention (CALI)—The CALI arm consisted of seven computerized instructional modules selected from the Earobics software (Cognitive Concepts, 2000a, 2000b) and Laureate Learning software (Semel, 2000; Wilson & Fox, 1997). Like the FFW-L exercises, the CALI modules targeted discrimination and memory of nonspeech sounds (viz., Earobics Karloon's Balloons), detection of individual phoneme changes (viz., Earobics C.C. Coal Car), phoneme discrimination (viz., Earobics Paint by Penguin), identifying matched syllable pairs (viz., Earobics Hippo Hoops), discriminating between minimal pair words (viz., Earobics Duck Luck), recalling commands (viz., Laureate Learning Systems Following Directions Series), and comprehending grammatical morphemes and complex sentences structures (Laureate Learning Systems Micro-LADS 1–6). None of the auditory stimuli in any of the CALI modules were modified acoustically. Similar to the FFW-L condition, each of the modules in the CALI arm were presented for 20 minute periods (a total of 1 hour and 40 minutes per day), five days per week for six weeks under headphones in a quiet environment. Like the FFW-L treatment arm, a certified and licensed speech-language pathologist supervised every session and monitors, who were graduate students in speech-language pathology, observed the students and provided positive nonverbal feedback. Children started at the beginning of each computer game in CALI. When a child attained 90% correct at a particular level for two days in a row, the level of the exercise was considered to be mastered, and the next level was presented. If a child reached mastery level on a game before the end of the six-week intervention session, he or she repeated that game until the last day of the intervention session.

Individual Language Intervention (ILI)—ILI was delivered by certified and licensed speech-language pathologists (SLPs) who worked with each child in a quiet room. Like the FFW-L and CALI conditions, each of the intervention sessions in the ILI condition lasted 1 hour and 40 minutes per day, five days per week, for six weeks. Our approach to literature-based language intervention was influenced by Gillam and Ukrainetz (2005), Norris (1993; 1989), and Strong and Hoggan (1996), who have suggested that language therapy with school-age children should target specific language skills within activities that are related to the topic and content of children's books. Opportunities for functional and interactive exchanges between clinicians and children (Gillam & Ukrainetz, 2005; Leonard, 1999; Nelson, Camarata, Welsh, & Butkovsky, 1996; Paul, 2002) were an integral part of therapeutic interactions.

The ILI activities were designed to target semantics, syntax (morphosyntax and clause structure), narration and phonological awareness. The ILI units were developed around 13 picture books (Appendix A) that were interesting to school age children, that could be read in a short amount of time, and that contained a variety of vocabulary words that ranged in difficulty level for children who were six to nine years of age. A minimum of six book units, usually one per week, were used with each child over the six-week course of intervention. The SLP chose the book unit that she thought would be the most interesting for the child with whom she worked. Appendix B presents the general outline that was used as the standard for each book unit.

Language intervention activities were developed for three levels of difficulty within each of the target areas (i.e., semantics, grammatical morphology, clause structure, narration and phonological processing). Appendix C presents the language targets within the three levels of difficulty in each of the five areas.

Throughout the ILI sessions, the SLP used a variety of language facilitation strategies that have been shown to be effective with children. These strategies included slower rate (Weismer, 1997), emphatic stress on target forms (Weismer, 1997), growth-relevant recasts (Camarata, Nelson, & Camarata, 1994; Nelson et al., 1996), focused stimulation (Cleave & Fey, 1997; Fey et al., 1993), incidental teaching (Kaiser, Yoder, & Keetz, 1992), scaffolding (Schneider & Watkins, 1996), and mediation (Miller, Gillam, & Peña, 2001).

The SLPs used a system developed by Miller, Gillam, and Peña (2001) to track the children's level of performance each day. After each activity, the clinicians judged the amount of teaching effort (low, medium, high) and the amount of student responsiveness (low, medium, high). Students advanced to the next level of difficulty when clinicians rated their teaching effort as low or moderate and rated student responsiveness as moderate or high for activities that were repeated in two consecutive sessions.

Academic Enrichment (AE)—The AE arm was designed to serve as an “active” comparison arm (Herbert & Gaudiano, 2005; Lohr, MeMaio, & McGlynn, 2003). AE shared common features with FFWL and CALI such as intervention setting, amount of contact with computers, type of child involvement, time spent on different computer games, reward system and type of clinician involvement. AE was also designed to be a nonspecific intervention comparison. Even though many of the computer games in AE contained language, none of the software games were specifically designed to promote the development of language or auditory processing skills. Rather, the computer games in the AE condition had been designed to teach mathematics, science and geography. The computer games included: Magic School Bus Discovers Flight (Scholastic, 2001a), Magic School Bus Explores Dolphins and Whales (Scholastic, 2001b), Coin Critter (Nordic Software, 1999), Zurk's Rainforest Adventure (Soleil Software, 1998), My Amazing World Explorer (Dorling Kindersley, 1999), Dinosaur 3D (Knowledge Adventure, 1999), and selected games from Arthur's 1st grade (TLC Educational Properties LLC, 1999a) and Arthur's 2nd grade (TLC Educational Properties LLC, 1999b). None of these games targeted language or reading directly.

General Procedures

Each child attended the summer intervention program for three-and-one-half hours daily. Of this time, one hour and forty minutes were devoted to the intervention conditions. Children also received a 20-minute snack break and attended a one-and-one-half hour group activity period in which they played board games, worked on informal arts and crafts (coloring, painting, cutting, etc) and participated in general outdoor activities (recess). The activity groups were managed and monitored by a classroom teacher and two aides. Children were randomly assigned to one of two activity groups (A or B). The order of intervention and activity groups (first half or second half of the morning) was alternated each week.

Attendance, Compliance, and Treatment Fidelity

On rare occasions, an SLP in the ILI condition would be absent or technical problems would lead to missed computer sessions. When an SLP was ill or when a computer was not useable due to technology problems, the child assigned to that condition would miss that day of treatment. Children's average attendance rate overall was 28 of 30 days. Only three of 216 children withdrew from the study during the intervention period. There was no difference between the average number of days present for the FFW-L ($M=27.31$; $SD=3.30$), CALI ($M=27.80$; $SD=2.75$), ILI ($M=27.27$; $SD=3.30$), and AE ($M=28.31$; $SD=2.13$) arms [$F(3,212) = 1.231, p = .299$].

Clinicians documented compliance each day. Children received stickers after they completed each treatment session. At the end of the week, children selected a toy from a prize board if

they had earned five stickers. There was a 7-step procedure for managing noncompliance. If a child refused to engage in an activity, the clinician verbally encouraged the child to resume her or his participation. If the first step was not successful, the clinician would offer the child an opportunity to win an extra sticker. In the third step, children could take a short break and get a drink of water. If that was not successful, children were offered a small food reward, such as a snack. In the fifth step, children were given a short (5 minute) time out period. If children were still noncompliant after a time-out period, we ended the session for the day. We planned to met with the parents to discuss the child's involvement with the study (step seven) if a child reached step six for three days in a row. Table 2 displays the average number of days (out of 30) that any compliance step was recorded in the therapy logs. Verbal encouragement was usually sufficient to obtain compliance. A multivariate comparison of the number of treatment days that compliance steps were recorded yielded no significant group differences. In addition, there were no important adverse events or side effects for participants in any of the arms of the study.

Clinicians were trained to administer the interventions. No clinicians provided treatment in more than one arm. The clinicians documented the number of games that the participants played (computer arms) or the number and type of activities that were undertaken (ILI arm) each day. The research team assessed treatment fidelity each week during the summer program. The sessions in all four conditions were videotaped each day. At the end of each week, a randomly selected videotape of a session in each treatment condition was sent to another site for review. Research assistants reviewed the use of positive reinforcement, the presence of a quiet and non-distracting environment, use of facilitative talk (in the ILI units), wearing of headphones (for the computer conditions) and the extent and type of clinician assistance. The results of the fidelity reviews were faxed to the site coordinators each Monday. Any modifications or suggestions resulting from the fidelity review were implemented at the intervention site.

Measurements

Children received a battery of language, literacy, and auditory processing measures at four different times throughout the course of the study: before treatment, immediately after treatment (i.e., 6 weeks after pre-testing), 3 months after treatment and 6 months after treatment. This paper focuses primarily on the results of the primary outcome measure (the *Comprehensive Assessment of Spoken Language, CASL*; Carrow-Woolfolk, 1999) and an auditory processing measure (Backward Masking). Other articles will focus on the secondary measures. However, the results of some secondary measures that inform the interpretation of the primary outcome variables are reported.

Comprehensive Assessment of Spoken Language—The primary outcome measure was the core composite score on The *Comprehensive Assessment of Spoken Language (CASL)*; Carrow-Woolfolk, 1999). The CASL is a norm-referenced, standardized test of expressive and receptive language. It was selected because it had strong psychometric properties and measured a wide range of receptive and expressive language skills. The core composite for children between the ages of 7;0 and 9;11 was based on the sum of the standard scores from the Antonyms, Syntax Construction, Paragraph Comprehension, Nonliteral Language, and Pragmatic Judgment subtests. Children between the ages of 6;0 and 6;11 years of age received the Nonliteral Language subtest, but scores on this test were not used to compute the CASL Core Composite. Like all global language tests, high correlations among the subtests and between the subtests and the composite score suggest that the test measures one general language skill.

Backward Masking—Backward masking is widely considered to be a measure of temporal resolution because it requires listeners to distinguish between sounds in time (Hill, Hogben, &

Bishop, 2005; Marler, Chaplin & Gillam, 2001). As with all other psychoacoustic measures, attention and memory contribute to performance on backward masking.

The backward masking task was delivered via a computer according to procedures that have been summarized in Marler, Champlin, and Gillam (2001), Thibodeau, Friel-Patti, and Britt (2002) and Wright, Lombardino, King, Leonary, Puranik, and Merzenich, et al. (1997). The task was administered in a quiet room with an ambient noise level of 45 dBA or less. Children looked at a computer screen showing three colorful pictures as they listened to three auditory intervals presented over TDH 39 headphones. One auditory interval contained a signal (a tone) that was closely followed by a masking noise. The other two auditory intervals consisted of the masking noise presented without signals. The child's task was to point to the picture (first, second, or third) that corresponded to the interval that was different, (i.e., the one that contained both the tone and the masking sound.). This is commonly referred to as a three interval, three alternative, forced-choice paradigm.

The signal was 1-kHz with an overall duration of 20 ms with a 10-ms envelope. The starting phase of the signal was fixed at zero degrees. The onset of the signal preceded the onset of the masker by 20 ms (the duration of the signal). The masker was a narrow-band noise that extended from 0.6 kHz to 1.4 kHz. Beyond the two cutoff frequencies, the masker level decreased at the rate of -96 dB/octave. The overall duration of the masker was 300 ms with a 10ms envelope. The pressure spectrum level of the masker was about 42 dB/Hz (overall level = 71 dB SPL). The signal and masker were randomly assigned to one of the three intervals; the other intervals contained the masker alone.

A training session was conducted to familiarize each child with the test environment and procedures. Children who were seated at a computer looked at colorful pictures, one corresponding to each of the listening intervals that were displayed on the computer screen. The examiner explained that there were three sounds. After listening carefully, the child was instructed to indicate the sound that was different from the other two by clicking the mouse pointer on the appropriate picture. A colorful picture would appear on the screen after each correct response, but a solid horizontal green bar would appear after each incorrect response.

Following the verbal instructions, the examiner provided a demonstration trial via a small loudspeaker. The signal was randomly presented in one of the three observation intervals. The observation intervals were 350 ms in duration and were separated by 500-ms pauses. The examiner indicated the interval that contained the signal by selecting one of the buttons. The child was given the opportunity to perform 5 practice trials independently. The headphones were then positioned over the child's ears, and 5 practice trials were performed.

During testing, the examiner was seated behind the listener and only assisted to redirect attention to the task when necessary. When ready, the child selected one of three buttons to begin the test. Three listening intervals were presented, and the child chose the one s/he thought contained the signal. If the response was incorrect, the first trial was repeated, systematically increasing the signal level, until the child responded correctly. After a correct response was obtained on the first trial, the signal levels for all subsequent trials were automatically adjusted based on the listener's response. The signal became louder (easier to hear) after each incorrect response and softer (more difficult to hear) after two consecutive correct responses. The initial step size was 5 dB, and the final step size was 2 dB.

A test run consisted of up to 60 trials and took 3 to 5 minutes to complete. The threshold estimate was based on the average of the last 10 of 12 reversals within a run. Threshold variability was based on the standard deviation from the same 10 reversals. If the threshold variability was greater than twice the step size, the run was repeated. A maximum of three runs may have been attempted per test. The first run to satisfy the variability criterion was taken as

the estimate of threshold, unless three runs were needed after which the average of the three runs was taken. The child received verbal encouragement and a tangible reward upon successful completion of each run. After the first threshold estimate, the child was given a break and returned later the same day for a second measure. The first threshold estimate was considered practice and the only the second threshold estimate was used in the statistical analysis. The extensive training and practice that preceded the second threshold estimate were designed to minimize practice effects, which have been demonstrated in previous studies (Bishop, Carlyon, Deeks, & Bishop, 1999; Wright, 2001).

Secondary Measures—The FFW-L and CALI arms primarily targeted sentence comprehension and phonological processing. These abilities were not specifically assessed by the subtests on the CASL or by the backward masking measure. We administered additional secondary measures that related more specifically to additional skills that were targeted in the FFW-L and CALI arms. The *Token Test for Children* (Disimoni, 1978) was administered to assess sentence comprehension. Due to the age and ability levels of the children in this study, there were ceiling effects on the first two sections of the Token Test. We report the combined raw score for sections three, four, and five in which children manipulated small tiles in response to examiner directions such as, “Touch the red circle and the yellow square” (Part III); “Touch the small blue circle and the small red circle” (Part IV); and “Put the green square beside the yellow circle” (Part V).

We administered the Blending Words subtest of the *Comprehensive Test of Phonological Processing* (Wagner, Torgesen, & Rashotte, 1999) to assess phonological awareness. This subset required children to listen to audiorecorded sounds, presented one at a time, in sequence. Children combined the sounds mentally and said the complete word (e.g., “What word do these sounds make: j-u-m-p? Children answer, “jump”).

The primary and secondary measures were administered in daylong testing sessions that included snacks, craft breaks, and lunch. If a child was fatigued, we discontinued testing for a short period of time. There was no particular order for test administration. Once administered, language and literacy measures were re-scored and checked for accuracy by three independent scorers.

Blinding

The principal investigators, data collectors, data analyzers, and members of a data safety and monitoring committee were blinded to the randomized treatment assignments. A research coordinator at each site was the only person who had a record of subject names, subject identification codes, and treatment assignments.¹

Parents were not told which treatment their children received, and they were not allowed to observe therapy. It is possible that children told their parents enough about their experiences at school to enable parents to deduce the arm they were randomized to. To assess this possibility, 73 parents completed a questionnaire either in written format or via phone survey in which they were asked to guess which treatment their child received. Only 20% of parents identified the correct arm that their child was randomized to, which was slightly less than chance. These results suggest that the majority of the parents were blind to the treatment that

¹There was a compromise to blinding at one site when some therapy files containing identifying information were reviewed. This occurred after the year 2 posttest data had been analyzed and reported to the Data Management Center at the University of Iowa, but before the year 3 pretesting had begun. The primary and secondary outcome data were not observed or analyzed, and no person involved with testing CLIP participants, scoring tests, analyzing test results, or reporting data to the Data Management Center had any knowledge of the compromise to blinding. The individuals involved had no subsequent involvement in data collection or analysis. The incident was reported to the university's Institutional Review Board and the NIH Data Safety Monitoring Board. Both committees determined that there were no adverse effects from this event.

their children received. It is unlikely that potential parental expectations of the value of a particular treatment influenced the outcomes of the study.

Baseline Comparison

To check for selection bias, we tested for potential pre-intervention differences between the participants in the four arms (Berger, 2005). Cochran-Mantel-Haenszel tests were used to assess categorical variables, stratified by site and year. For continuous outcomes or covariates, mixed models Analysis of Variance (ANOVA) were used. No statistically significant differences were found among the four arms for sex, age, socio-economic status or ethnicity. The mixed model ANOVAs (with site, sub-site, and classroom as a random effects) were conducted to examine whether the three intervention and comparison groups differed at baseline on their pre-test scores for each outcome. No statistically significant differences among the four arms were identified.

Attrition

There were three participants who only received pretesting because they withdrew from the study prior to completing the treatment. Two additional participants missed two of the four testing sessions and eight participants missed one testing session (a total of 21 missed testing sessions of 648 possible sessions; 2.3% missing data). Preliminary analyses with mixed model ANOVAs (with site, sub-site, and classroom as a random effects) were conducted to compare pre-test scores for participants who had completed the post-test versus participants with pretest scores who did not complete one or more post-tests. Preliminary results indicated no significant differences in the complete cases versus incomplete cases on pre-test scores.

Results

The overall goal of this RCT was to compare the language and auditory processing outcomes of three types of language intervention and an active, nonspecific language intervention comparison group. If the temporal processing deficit theory of language impairment was correct, it was hypothesized that the participants who received the FFW-L intervention would show greater improvement on a standardized language test than children who were assigned to the other treatment arms or to the active, nonspecific comparison group at immediate posttest and at 3 month and 6 month follow-up periods. A Mixed Model Analysis of Covariance (ANCOVA) was used to test the original hypotheses. The primary dependent variable was the CASL core composite score. Performance on the CASL composite pre-test was significantly related to the post-test and follow-up outcomes, $F(1,207) = 488.6, p < .001$; adjusted R-squared = .62. Because of this positive and significant correlation, the pre-test score on the CASL composite score was used as a covariate. Including the pre-test as a predictor in the model effectively adjusted the post-test and follow-up tests for differences on the pre-test scores. The fixed independent variables were the four arms (AE, CALI, FFW-L, and ILI) and three post-intervention times (post-test, 3-month follow-up and 6-month follow-up).

The Mixed Model utilized a multilevel model approach to avoid the possibility of statistically significant differences due to the decreased standard error resulting from non-independence in participant scores within a particular sex, year (1, 2, or 3), site (Austin, Dallas, or Lawrence), sub-site (particular school), or classroom group (A or B) (Bryk & Raudenbush, 2002; Snijders & Bosker, 1999). If the variance in participant scores attributable to sex, year, site, or classroom was zero, that variance component was dropped from the model (Kreft & De Leeuw, 1998). The degrees of freedom for statistical significance tests were calculated using the Kenward-Roger adjustment for degrees of freedom in mixed models for repeated measures, because it reduces the potential for Type I error (Littell, 2002). For the within-subjects effects, the unstructured variance-covariance matrix was based on comparing seven different structures

using the Bayesian Information Criterion. An effect size was calculated using a pseudo R-squared based on the Level 1 residuals proposed by Snijders & Bosker (1999), which examined the proportional reduction in the residual variance. Standardized residuals from the hypothesis models were examined to identify any outliers and ensure that unusual values were accurate. Table 3 presents the number of cases (N), means and standard deviations (SD) for the CASL composite score and the backward masking score by Treatment and Time. Across time, there was a positive trend for the CASL composite score and a negative trend for the backward masking, meaning that performance on both tasks improved.

Pre-post Comparisons

Because the statistical model that specifically addresses the hypotheses and research questions does not directly test whether change occurred from pretest to posttest, a preliminary mixed model repeated-measures analyses was performed to examine this issue of clinical importance. The predictors were Arm (CALI, FFW-L, ILI, AE) and Time (pre and post-test) and the Arm x Time interaction. The dependent variable was the CASL core composite standard score. The analysis used a mixed model methodology that incorporated site and sub-site random effects and an unstructured variance-covariance matrix for the within-subject residuals. There was a significant Time effect, $F(1,212) = 97.9$, $p < .001$; $d = 1.08$, indicating significant improvement from pre- to posttest across the four arms of the study. Neither the Arm main effect nor the Time x Arm interaction was significant.

Hypothesis Tests

The primary research question was: Will participants randomly assigned to the FFW-L arm show greater improvement on the CASL composite score than the participants assigned to the treatment arms (CALI or ILI) or to the active comparison (AE) arm immediately after intervention and at 3-and six-month follow-up periods? As discussed earlier, performance on the CASL composite pre-test was significantly related to the post-test and follow-up outcomes. The Arm main effect was not significant, $F(3,206) = 1.9$, ns, indicating that the children in one arm of the study did not benefit from their experiences to a greater extent than the children in any other arm. The Time main effect was significant, $F(2,211) = 43.3$, $p \leq .001$, adjusted R-squared = .05, indicating that the children in all four arms demonstrated statistically significant change on the CASL from posttest to the three and six-month follow-up periods. The least squares estimated adjusted means across time were: 83.0 at post-test, 85.0 for the 3-month follow-up, and 86.8 for the 6-month follow-up. All three means were significantly different from each other using Tukey-Kramer adjusted p-values ($p < .001$). The Arm by Time interaction testing differential change across groups and time was not significant, $F(6,280) = 0.8$, ns, indicating that the lack of differences among study arms did not vary across time.

It is possible that the core composite standard score on the CASL masked group differences among the language skills that were measured by the CASL subtests or that there were group differences in language abilities that were not measured by the CASL subtests. Table 4 presents the unadjusted means and standard deviations for raw scores on four subtests on the CASL (antonyms, syntactic construction, paragraph comprehension, and pragmatic judgment). The raw score means and standard deviations for the sum of Parts III, IV, and V of the *Token Test for Children* (Disimoni, 1978), and the raw scores for the Blending Words subtest on the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner, Torgesen, and Rashotte, 1999) are also presented. Similar to the CASL analysis, mixed model ANCOVA were performed on each measure with Arm (CALI, FFW-L, ILI and AE) as the between factor and Time (immediate posttest, 3-month follow-up, and six-month follow-up) as the within factor. The pretest score for each measure was entered as the covariate. The results for each CASL subtest and the Token Test for Children were consistent with the results for the CASL core

composite: the Time main effects were significant, indicating change over time, but the Time x Arm interactions were not significant.

The blending words subtest of the C-TOPP yielded a significant Time main effect [$F(2,422) = 7.214, p < .01, \eta^2 = .034$], and a significant Time x Arm interaction [$F(6,422) = 2.606, p < .05, \eta^2 = .036$]. The means for the three assessment times differed significantly (Tukey, $p < .001$). At posttest, the means for the CALI, FFW-L, and ILI arms were above the 95% confidence interval of the mean for the AE arm, indicating that the children in the three treatment arms outperformed the children in the active comparison arm on a measure of phonological awareness. There were no significant differences between the arms at the 3-month follow-up. However, at the 6-month follow-up, the means for the CALI and the FFW-L arms exceeded the upper boundaries of the 95% confidence intervals for the AE and ILI arms. Therefore, it appears that the intervention in the CALI and FFW-L arms was better than the intervention in the AE and ILI arms for improving phonological awareness.

Recall that the primary hypothesis (that children assigned to the FFW-L arm would have better outcomes than children in the other three arms) was based on the temporal processing deficit hypothesis (Tallal, 2004). Results of the language testing were not consistent with the hypothesis. It is possible that the scores on the language measure were not reflective of changes in auditory temporal processing. Therefore, we applied the same statistical model using backward masking as the dependent variable. If the acoustic modification that was built into the FFW-L software ameliorated a temporal processing deficit, the backward masking scores of the children in the FFW-L arm should have improved more than the backward masking scores of the children in the other three arms, which did not include modified acoustic input.

The pre-test backward masking score was significantly related to the post-test and follow-up outcomes, $F(1,193) = 214.9, p < .001$; adjusted R-squared = .38. When the posttreatment scores were adjusted to reflect differences in pretreatment performance levels, the Arm main effect was not significant, indicating no differences between the three treatments and the active comparison arm. The Time main effect was significant, $F(2,211) = 20.9, p < .001$, adjusted R-squared = .04, indicating that the children in all four arms demonstrated meaningful change on backward masking from posttest to the three and six-month follow-up periods. The least squares estimated adjusted means across Time were: 55.7 dB at post-test, 53.5 dB for the 3-month follow-up, and 49.6 dB for the 6-month follow-up.

Another model tested the three-way interaction between backward masking, Time and Arm to determine if there were group differences depending on the initial backward masking score. This interaction was not statistically significant. The results of these analyses suggest that children's auditory temporal processing skills improved whether the treatment they received contained modified speech or not.

We also conducted an analysis to examine whether the FFW-L group or other groups differed on the CASL composite score depending on both their backward masking score and time. The 3-way interaction between arm, backward masking score, and time, and the 2-way interaction between arm and backward masking score were statistically nonsignificant. Participants with lower backward masking scores did not show greater language skills improvement in the FFW-L condition compared to the other arms.

Clinical Significance

Statistical analyses suggest that the children in all four arms made similar amounts of improvement on a global test of receptive and expressive language. Because this was an intervention study, we asked whether the changes that were demonstrated by the participants were clinically relevant. We used effect sizes, confidence intervals, and normalization

percentages to assess the clinical significance of the results at posttest and at the six-month follow-up testing.

The effect sizes for the intention to treat analysis and for the raw score changes on backward masking, the *Token Test for Children* (DiSimoni, 1978) and the Blending Words subtest on the C-TOPP (Wagner, Torgesen, & Rashotte, 1999) are presented in Table 5. There are moderate CASL effect sizes from pretest to posttest across arms as reflected by Cohen's *d* values that ranged from .56 to .79. The gains from pre-test to the six-month follow-up test are uniformly large as reflected by Cohen's *d* values that ranged from .93 to 1.34. The children in the CALI group obtained moderately-large effect sizes on the backward masking measures from pretest to posttest and large effect sizes from pretest to six-month follow-up testing. The children in the other three arms evidenced moderately-small to moderate effect sizes on backward masking immediately after treatment and moderately-large effect sizes from pretest to six-month follow-up. The effect sizes for performance on the Token Test were unremarkable at immediate posttest. At the six-month follow-up period, the effect sizes for the children in the CALI, FFW-L, and ILI arms were in the moderate range (.54 – .66), but there were relatively small effects (.35) for children in the AE arm. The effect sizes for the Blending Words measure were noticeably different across the arms of the study. Immediately after treatment, the participants in the AE arm evidenced a negative effect (a decrease in performance) while the children in the other three arms evidenced a moderately-small effect. At the six-month follow-up testing, the Cohen's *d* effect sizes for participants in the FFW-L arm were moderately high while the effect sizes for the children in the other three arms were in the moderate range. For this measure of phonological awareness, the effect size for the FFW-L group was 27% higher than the effect sizes for children in the CALI arm (.79 vs .62), 49% higher than the children in the ILI arm (.79 vs .53) and 72% higher than the children in the AE arm (.79 vs .46).

Another indicator of clinical significance is the participants in each arm whose posttest and six-month follow-up scores were outside the upper limits of the 68% and 95% confidence intervals (1 SEM and 1.96 SEM) for the CASL Core Composite Score. As shown in Table 6, over half of the children in each arm earned posttest scores that were above the upper limits of the 68% confidence interval immediately after the intervention period. Between 63% and 80% of the children in each arm earned six-month follow-up scores that were greater than one standard error of measurement on the CASL Core Composite. The 95% confidence interval is a more conservative indication of clinically significant change. More than half of the children in each arm earned scores that were above the upper limit of the 95% confidence interval for their pretest score by the six-month follow-up testing period.

The final indicator of clinical significance was the percentage of the children in each arm whose scores moved into the normal range on the CASL over the course of the study. Following Tomblin, Zhang, Buckwalter & O'Brien (2003), the normalization criterion was a standard score of 83 or above. According to this criterion, the percentages of children who normalized over the course of the study were 64.8% for CALI, 59.3% for FFW-L, 75.9% for ILI, and 70.4% for AE.

Discussion

The primary purpose of this randomized controlled trial was to determine whether Fast ForWord – Language, a computer instructional program designed to improve auditory temporal processing skills in children with language impairments, was more effective than other types of interventions for improving language and auditory processing. If language impairments primarily reflect a deficit in auditory temporal processing skills, children randomized to the FFW-L arm should have presented greater pretest to posttest improvement on a standardized measure of language and on a measure of backward masking in comparison to children in the

other conditions. Further, if the auditory processing deficit was remediated, the FFW-L advantage should have been maintained over time.

The children with language impairments in our sample who received FFW-L did not fare better than children in other language interventions of equal intensity on our primary outcome measure of language, the CASL core composite score, on the individual CASL subtests, on the *Token Test for Children* (DiSimoni, 1978), or on a measure of temporal auditory processing, the backward masking threshold. With one exception, the children in all four arms made significant improvement on the language and the auditory processing measures from pre-to post-testing as well as the follow-up testing three and six months later. The exception was the Blending Words subtest of the C-TOPP (Wagner, Torgesen, & Rashotte, 1999), which is a measure of phonological awareness. On that measure, children in the CALI and FFW-L arms evidenced significantly better performance than the children in the AE and ILI arms.

One important finding is that the primary measures of language and auditory processing improved significantly across groups. One possible conclusion from this pattern of results is that all four arms were equally ineffective. That is, the gains in language and auditory processing that were observed across the four arms of the study may have arisen as a result of practice effects, maturation, regression to the mean, and/or spontaneous recovery. If we had employed a no-treatment comparison group in the design of the study, an interpretation of equally ineffective intervention practices would be entirely appropriate. However, this study was designed with an active comparison group who received daily computer assisted instruction on academic subjects. Without a no-treatment control group, we have no way to assess the relative extent to which intervention and changes over time contributed to changes in the primary outcome variables. Most of the children in this study had moderate-to-severe language impairments. We felt an obligation to their parents to provide each participant with some sort of positive experience. By including an active control group and an ILI group, we controlled the variables (attention by an interested clinician, engagement in activities designed to promote thinking and problem-solving, and opportunities to socialize with peers) that would have been uncontrolled if we had included a “no treatment” group.

Language Measures

Careful consideration of the pattern of language results in light of what is known about the longitudinal course of language impairment supports the validity of the statistical results demonstrating significant improvements in language and auditory processing skills for participants in all four arms. A number of longitudinal studies have shown that a high percentage of school-age children with language impairments do not make gains beyond one standard error of measurement on formal tests administered one, three, or even as many as 10 years after they were identified (Aram & Hall, 1989; Bernstein & Stark, 1985; Conti-Ramsden & Botting, in press; 1999; Fazio et al., 1996; Silva et al., 1987; Snowling & Hulme, 1989; Tallal et al., 1996). In this study, the six-month follow-up scores of 63 to 80 percent of the participants were more than one standard error of measurement above their pretest scores. In addition, more than 50% of the children had six-month follow-up scores that were outside the 95% confidence interval of their pretest scores, yielding large pretest to posttest effect sizes across the groups (range of .93 to 1.34). These findings support the conclusion that children in all four arms made clinically relevant improvements on the CASL that are greater than would be expected for children with language impairments who have participated in longitudinal studies.

An additional measure of clinical significance adds more support to our conclusion of improved language skills across the treatment and comparison arms. At the time of the six-month follow-up evaluation, between 59 and 76 percent of the participants had good outcomes as indicated by CASL scores that were above Tomblin et al's (2003) suggested cut-off score of 83. In many

previous studies, between 55 and 80 percent of children with language impairments evidenced follow-up scores on global language tests that remained in the impaired range (Beitchman, Wilson, Brownlie, Walters, & Lancee, 1996; Conti-Ramsden & Botting, 1999; Tomblin, et al., 2003). It is important to remember that in the current study, children were identified as language impaired with one measure (*The Test of Language Development-Primary: 3rd Edition*; Newcomer & Hammill, 1997) that differed from the primary outcome measure (the CASL). This procedure was designed to control for regression effects by minimizing regression to the mean after pretesting (Tomblin et al., 2003). In comparison to other studies of similar age children with similar types and degrees of language impairments, the treatments used in this study reduced the persistence of language impairment by half or more.

The children made less improvement on a test of sentence comprehension than they did on the primary outcome measure, which was a global language test. The immediate posttest and six-month follow-up effect sizes for the *Token Test for Children* (DiSimoni, 1978) were approximately half as large as the effect sizes for the *Comprehensive Assessment of Spoken Language* (Carrow-Woolfolk, 1999). Performance on all the subtests on the CASL requires some degree of language comprehension. But, only one subtest, Paragraph Comprehension, primarily assesses language comprehension. The rest of the subtests of the CASL involved language production. The difference in the effect sizes of the gains on the CASL and the *Token Test for Children* suggests that children's expressive language skills improved more than their receptive language skills. This occurred despite the fact that three of the four arms of the study primarily focused on listening. These results are consistent with the outcomes of a recent study that compared computerized intervention using modified or unmodified speech (Bishop, Adams, & Rosen, 2006). In that investigation, the children in both intervention groups and a no-treatment control group made relatively small gains on a test of sentence comprehension.

There was one language measure that yielded significant group differences. At the immediate posttest, participants in the FFW-L and CALI arms earned higher scores on the Blending Words subtest of the *Comprehensive Test of Phonological Processing* (Wagner, Torgesen, & Rashotte, 1999), than the participants in the AE arm. At the six-month follow-up, scores on the blending words subtest for participants in the FFW-L and CALI arms were still significantly higher than the scores for participants in the AE and ILI arms. It appears that the two computerized instructional programs that focused primarily on auditory discrimination of sounds, syllables, and words yielded better phonological awareness results than a computerized treatment that focused on general academic skills or a clinician-directed language treatment that emphasized phonological awareness to the same extent as other language skills (vocabulary, syntax, grammatical morphology and narration). Finally, similar positive outcomes for phonological awareness for the FFW-L and CALI arms suggests that the inclusion of modified speech in the FFW-L program does not appear to be necessary for improvements in phonological awareness.

Auditory Processing

We used a backward masking task to assess aspects of temporal processing. In backward masking, a brief tonal signal precedes a long-duration noise masker. We patterned this listening task after one developed by Wright et al. (1997), who reported deficits in temporal resolution in children with language impairment. Like Wright et al., we found that some children with language impairment had difficulty distinguishing the signal from the masker that followed. However, there were some children with language impairments who performed relatively well on the backward masking measure at pretest. In our study, children in all four groups evidenced improvement on the backward masking measure over time regardless of the level of their pretest performance. These results suggest that individual instruction that involves intensive daily experiences in listening and responding to verbal input had a positive effect on the ability to

perform a backward masking task, whether that intervention was delivered by a human agent or by a computer software program.

What mechanism might be responsible for deficits and improvement on our temporal processing measure? There appear to be two schools of thought on the topic. One view is that difficulties with temporal processing occur because the incoming stream of information is altered so that fluctuations in the waveform are smeared in the time domain. If this were the case, a child with language impairment would have difficulty encoding auditory input, which would yield a poor quality template for learning language. While there is no direct evidence that this actually occurs, there are physiologic data that indicate auditory brainstem responses (ABRs) recorded from children with specific language impairment (Marler & Champlin, 2005) or more generalized learning impairment (e.g., Kraus et al., 1996) are significantly different from those obtained from children with typically developing language who were matched by age and non-verbal intelligence. Because the ABR is not influenced by the participant's state of consciousness, cognitive factors such as attention, memory, motivation, etc. are not believed to affect the magnitude of this obligatory physiologic response. Therefore, some of the children in the present study may have been afflicted by suboptimal neural encoding, and their participation in generalized listening experiences may have changed the functional anatomy and physiology of brain areas that were used for creating auditory templates.

Alternatively, it is possible that general cognitive abilities related to attention, memory, and perception play a similar role in detecting a signal in the presence of backward masking. Children with language impairments may have generalized information processing difficulties (Bishop, Adams, & Rosen, 2006; Ellis Weismer & Evans, 2002; Hoffman & Gillam, 2004; Johnston, 2006). If this were the case, then detecting the signal in the presence of backward masking would be a relatively difficult task for a number of reasons. It has been suggested that with appropriate training, signal thresholds in backward masking can be lowered. For example, Marler et al. (2001) reported that computer-based training with FFW-L over a five-week period yielded progressively lower signal thresholds in a small group of children with language impairment. Additionally, providing a brief sound in the opposite ear to cue the presence of the signal has long been known to reduce backward masking in adults (e.g., Puelo & Pastore, 1980). Introducing a contralateral marker may have helped children learn how to listen for the signal, which may have led to lower thresholds in backward masking. It is reasonable to suspect that both sensory and non-sensory factors contribute to a person's temporal processing ability. The three treatment arms and the active comparison arm all impacted the cognitive functions related to the information processing skills required for simultaneously listening to verbal information and looking at related visual stimuli for long periods of time. Even though children with language impairments are likely to represent a heterogeneous population in terms of the source and the magnitude of the cognitive processing deficits that contribute to their language impairments, it appears that most of these children can profit from instructional programs that stimulate multiple cognitive functions.

Critical Components of Language Intervention

Fast-ForWord-Language was designed to capitalize on the cerebral cortex's demonstrated ability to reorganize in response to focused, intensive stimulation. The core principles of the FFW-L program included heavy repetition, intensive intervention across successive days (one hour and forty-minute sessions, five days per week for four to six weeks), active attention to each trial, motivation through behavioral reinforcement, immediate feedback about response correctness, and programs that adapt to increased skill by increasing difficulty as a function of improvement (Agocs et al., 2006; Tallal, 2004). These principles were based on studies of training paradigms that led to neural change in animal models (Merzenich et al., 1996). The

outcomes of this RCT inform our understanding of the necessity of the core principles of neuroplasticity-based training for language intervention with children. Many of the games in the AE arm were not adaptive, involved minimal repetition, and offered no feedback on accuracy. Because the children in the AE arm had language outcomes that were very similar to those of the children in the FFW-L arm, some basic behavioral teaching principles that lead to behavioral and neurophysiological change in animal models may not be necessary components of protocols for teaching language to human beings. We did not obtain objective measures of neurophysiological change. However, a clear relationship between brain and behavioral changes has been established for children who received FFW-L training (Tallal & Gaab, 2006). It can be reasonably assumed that any kind of teaching or intervention leads to neural reorganization. Adaptive responses, repetition and feedback on accuracy probably play a role in human learning, but their roles may be mitigated by higher cognitive and linguistic functions.

Modified speech stimuli—The publishers of FFW-L have claimed that an additional important component of this instructional software is the adaptive modification of auditory stimuli that promotes rapid auditory processing (Tallal et al., 1996; Merzenich et al., 1996). We wanted to know if modified speech stimuli was better for promoting language development in children with language impairment compared to unmodified speech (as in our CALI condition). Consistent with earlier findings by Gillam et al., 2001, our results indicated that the modified speech in FFW-L was not a necessary component of interventions that positively affect immediate or long-term language or auditory processing outcomes.

Children in the FFW-L and CALI arms had better outcomes on a measure of phonological awareness than children in the AE and ILI arms. We did much to equate FFW and CALI, but there were many factors other than modified speech that distinguished these programs. A strict comparison would have required us to use FFW stimuli exactly, but without using the algorithm for modifying the speech children hear. We were not able to make these changes in the FFW-L program. The exercises in our CALI condition had natural speech, but they had a number of other differences as well. We really compared two different commercially available packages, both of which were delivered via computer and both of which were very intensive. This made an interesting and important contrast, but it is not a contrast that was based solely on modified speech. The critical point is that a computer program that did not contain modified speech yielded the same improvements in phonological awareness as a program that did.

Intervention agent—We found that children who received one-to-one language intervention with a certified speech-language pathologist fared as well as children who interacted with computers during their intervention time. However, one should consider that the time investment of the SLP is greater than that of the computerized interventions. In the three computer arms, 6 children received the computerized interventions at one time, with 2–3 monitors to oversee the children's play and to encourage the children when needed. In contrast, in the ILI condition, children worked with SLPs in one-to-one sessions. It is possible that the ILI condition could be adjusted to a group setting, and we are currently conducting studies to evaluate this possibility.

With respect to the two computer interventions versus the human (ILI) intervention, we believe we designed our study to take advantage of the best that each modality had to offer. The computer-directed intervention provided many discrete trials in a limited period of time, and the computer control allowed stimuli to vary slightly along specified dimensions. On the other hand, human (clinician) interaction afforded opportunities for the clinician to respond directly to the child's intended meaning within the context of goal-directed social interaction. It does not appear that differences related to discrete-trial learning and goal-directed social interaction

played a key role in the outcomes of this study. Our results suggest that both types of intervention can lead to clinically relevant changes on a global language test.

Comparisons to Previous RCT's

Our results are consistent with the RCT of Cohen et al. (2005) with a smaller group of children who had severe mixed language impairment in that there was no additional benefit of FFWL compared to another computerized intervention without modified speech. However, it is important to note that the children in our RCT made larger gains on a global language test than the children in the Cohen et al. (2005) RCT. This may have occurred because the children in their study, as a group, had more severe language impairments than the children in our study.

Our results were not consistent with those of Pokorni et al. (2004). They did not find significant improvement in language skills following FFW-L intervention, yet we did. This difference in results may be due to participant characteristics as well as the statistical power of the studies. Their participants had reading disorders and almost half of the children were at-risk because of poverty. The inclusion of children at risk for language difficulties may have meant that the participants included children whose language weaknesses were due to reduced exposure to a language-rich input and/or reduced opportunities to engage in language-stimulating interactions, and not due to the kinds of underlying factors that lead children to have problems learning language in the presence of ample input and interactions that would normally foster language development. In addition, Pokorni et al. (2004) had 16–20 children in their intervention groups; whereas we had 54 children per intervention group. Such factors may have limited the benefits of FFWL intervention in their study. Future studies with children who have language impairments and who are at-risk because of poverty will shed light on this finding.

The AE Active Comparison Arm

A final issue is our finding that children who were randomly assigned to an active comparison arm made improvements on our language and auditory processing measures that were similar to the improvements by participants in the three treatment arms. Recall that we did not hypothesize that the children in the AE arm would make significant gains on our outcome measures. We are somewhat limited in our ability to answer why the AE comparison arm was effective for some children. The computer games that we used varied in the amount of oral language that was heard by the children. For example, in the computer game “Coin Critters,” children received very little verbal input. They primarily added coin values and heard “okay” or “oops” in response to a correct or incorrect answer. In contrast, computer games such as “Magic School Bus Discovers Flight” contained some areas in the program in which new vocabulary was introduced and defined while other areas of the same program had an art activity with little verbal input. Although we could attempt to discern which aspects of each software package may have facilitated language, the children were free to choose whichever parts of the academic enrichment games that they wanted to pursue each day. Thus, some children may have spent more times in the picture gallery of Zurk's Rainforest snapping pictures of animals; whereas others may have spent more time categorizing and hearing labels of rainforest animals.

As we evaluated the commonalities across the conditions, there appear to be three possible explanations for the language improvements that were seen in the children who were randomized to the AE arm. The first is related to our overall program. The children attended school for nearly four hours each day, of which only one hour and 40 minutes was spent in their assigned condition. During the remainder of the day, the children participated in crafts, recess, snack and other activities. There was no structured language intervention during this time, nor a trained professional targeting language skills. However, many parents reported that their child developed friendships in the classroom setting. Typically, children with language

impairments are the least likely to be selected as a friend in the classroom (Gertner, Rice, & Hadley, 1994). Because the children were grouped with other children with language impairments, it may have led to a situation where friendships could be made more easily in comparison to classrooms in which only one or two children have language impairments. As a result of new bonds, the children may have had a very socially oriented classroom environment that facilitated their language skills. We are not able to evaluate this hypothesis with data from our study, but further study of classrooms with children with language impairments over short periods of time, like that in this study, may lead to some important insights on the value of productive and mutually satisfactory social interactions among peers, particularly among and between children who have language impairments.

Another possible explanation for the positive outcomes for the children in AE arm may be related to improvements in general information processing skills. The AE computer games were not devoid of language. They contained words, sentences, and music that may have been facilitative to the children. Listening under headphones for one and a half hours per day for 30 days may have tuned the children's listening skills. In addition, the children were receiving visual stimulation via the computer games. Intensive and simultaneous auditory and visual experiences may have improved the children's visual organization skills. Hoffman and Gillam (2004) found that children with SLI had a smaller overall capacity for visual-verbal and visual-spatial information along with significant difficulty processing visual information when both verbal and spatial domains were activated in dual processing tasks. The possibility exists that children who participated in the AE arm benefited from intense practice with visual and auditory stimuli as provided by the academic computer games and, subsequently, their information processing systems were modified in a way that was beneficial for test taking skills. Perhaps the instruction in all four arms facilitated the ability to focus, execute, sustain, and shift attention, as well as the ability to encode information. If so, improved attention and information processing abilities might be shared across the treatments, including stimulation with academic enrichment software. Additional research is required to examine the role of attention in language impairments and the potential impact of general facilitation of attention and information processing abilities.

A third possibility is that the similar results for the active comparison arm and the three treatment arms reflected the value of careful attention. All the clinicians who worked in this research study, regardless of the arm they were assigned to, believed that they were participating in something important that was likely to be beneficial to children with language impairments. It is likely that this positive attitude was conveyed to parents during the various interactions they had with project staff. Staff members also made children feel special because they were involved in this study. It was apparent to us that parents expected good results from their children's involvement in this study and children were excited and happy about their participation. It is possible that the core variable in our results was not changes in children's attention regulation; rather it may have been having the right kind of attention paid to them.

Summary

This RCT was conducted to determine whether children with language impairments who received Fast ForWord – Language made more improvements on measures of language and auditory processing than children who received other types of intervention. Children who were randomized to the FFW-L arm did not present greater pretest to posttest improvement on a standardized measure of language and on a measure of backward masking in comparison to children who received computer assisted instruction that taught auditory processing, memory and language without modified speech stimuli (the CALI arm), children who received individual language intervention with a certified speech-language pathologist (the ILI arm), or children in a active comparison group who received academically oriented computer assisted

instruction that was not specifically designed to improve language, reading or auditory processing skills (the AE arm). The children in all four arms made similar amounts of improvement during intervention as measured by a global test of receptive and expressive language and a test of backward masking, and improvements on these measures continued for the next six months. Children in the FFW-L and CALI arms made greater improvements on a measure of phonological awareness than children in the ILI and AE arms. Even though these results were not consistent with the temporal auditory processing hypothesis, they do not mean that auditory processing skills are not important for language development. As Johnston (2006) has pointed out, language development requires listening, and temporal processing is necessarily part of listening to speech. But, mechanisms such as auditory selectivity and volitional mental focus may be as important for processing and comprehending speech as the specific auditory abilities that are related to temporal processing. The results of this study suggest that intensive language intervention experiences that require close attending and immediate responding to auditory and visual stimuli in combination with opportunities for socialization with same-ability peers and a great deal of positive attention from caring and interested adults should result in clinically relevant improvements in language and auditory processing skills in children with language impairments.

Acknowledgements

This research was funded by NIH grant U01 DC04560 awarded to the first three authors. It also was supported by resources from NIH Grant P30 HD02528, The Kansas Mental Retardation and Developmental Disabilities Research Center, BNCD, P30 DC005803 at the University of Kansas, and the Lillywhite endowment at Utah State University. We acknowledge the dedication and assistance of Emily Tobey, Lori Betourne, and Alicia Wanek. We also acknowledge the guidance of our Project Officer, Julia Gulya. We thank Steven Camarata, Chris Dollaghan, Judith Gravel, Fred Gruber, and Mark Espland who served on our Data Safety Monitoring Board. Jack Fletcher, Judith Johnston, and Karen Rascati, who served on our Advisory Committee, provided important suggestions throughout the course of the project. Diane Anderson and William Clarke, at the University of Iowa, oversaw the data management with great diligence. This study could not have been conducted without the very able assistance of numerous speech-language pathologists, graduate students and undergraduate students. We are grateful to the school districts that participated: Blue Valley Schools, Dallas Independent School District, Kansas City Kansas Public Schools, Lawrence Public Schools, Leander Independent School District, Pflugerville Independent School District, and Plano Independent School District. Last, but not least, we thank the children and their families for their long-term commitment to this study.

Appendix A: Book List for the ILI Units

Amazing Grace
 Mary Hoffman
 Dial Books for Young Readers (1991)
 How Do Dinosaurs Say Good Night?
 Jane Yolen
 Blue Sky Press (2000)
 Saturday Night at the Dinosaur Stomp
 Carol Diggory Shields, Scott Nash (Illustrator)
 Candlewick Press (1997)
 Little Grunt and the Big Egg
 Tomie dePaola
 Holiday House (1993)
 Tops and Bottoms

Janet Stevens
 Harcourt (1995)
 Tacky the Penguin
 Helen Lester
 Houghton Mifflin (1988)
 Strega Nona
 Tomie De Paola
 Simon & Schuster (1998)
 Doctor Me Di Cin
 Roberto Piumini
 Front Street / Lemniscaat (2001)
 Can I Have a Tyrannosaurus Rex, Dad? Can I? Please?
 Lois G. Grambling
 Bridgewater Paperback (2000)
 Miss Nelson Is Missing
 Harry Allard and James Marshall
 Houghton Mifflin (1977)
 Alexander and the Terrible, Horrible, No Good, Very Bad Day
 Judith Viorst
 Aladdin Paperbacks (1972)

Appendix B

Individual Language Intervention Literature-based Language Unit Outline

Day	Minimum Time (in minutes)	Activity
ONE	5	Prestory presentation: Semantic map
	5	Prestory presentation: Preparatory set
	10	Read entire story
	5	Narrative Activity: Post story presentation -Discussion Questions
	5	Semantic Activity: Add vocabulary targets to New Word Book
	8	Semantic Activity Choice
	8	Syntax Activity Choice
	8	Syntax Activity Choice
	15	Narrative Activity: Narrative retelling through pictography
	8	Phonological Awareness Choice
	8	Phonological Awareness Choice
	8	Semantics Activity Choice
TWO	5	Semantics Activity: Review target vocabulary in New Word Book
	10	Reread the entire story
	20	Narrative Activity: Create a retold book: dictation —use target vocabulary
	8	Semantics Activity Choice
	8	Syntax Activity Choice
	8	Syntax Activity Choice
	8	Phonological Awareness Choice
	8	Phonological Awareness Choice
	15	Narrative Activity: Audiotape of retelling
		If there is additional time: Repeat any activities that the child had difficulty with in previous day(s) or give child the choice of prior activities
THREE	10	Reread the entire story
	5	Semantics Activity: Review the target vocabulary in New Words Book
	5	Read the child's retelling

Day	Minimum Time (in minutes)	Activity
	15	Narrative Activity: Create a parallel book —include target vocabulary
	8	Semantics Activity Choice
	8	Syntax Activity Choice
	8	Syntax Activity Choice
	8	Phonological Awareness Choice
	8	Phonological Awareness Choice
	15	Narrative Activity: Audiotape of a parallel story
		If there is additional time, repeat any activities that the child had difficulty with in previous day(s) or give the child the choice of prior activities

Appendix C

Levels of Language Targets for the ILI Units

	Level 1	Level 2	Level 3
Narratives	1. Setting 2. Abbreviated Setting	1. Basic Episode 2. Added Elements	1. Multiple Episode 2. Complex Episode
Semantics	5 concrete nouns	5 concrete nouns (from level 1) + 4 verbs	4 verbs (from level 2) + 8 adjectives, nouns or adverbs
Syntax Grammatical-morphology	1. Copula “be” 2. Auxiliary “be”	1. Auxiliary inversion 2. Past tense modals	1. Wh-Questions with modals 2. Past tense modal inversion
Syntax-Clause Structure	1. Coordination “and” 2. Simple infinitive complements	1. Coordination “but, or, so” 2. Subordinating conjunctions	1. Subject relative clauses 2. Objective relative clauses
Phonological Awareness	1. Rhyming 2. Sound matching	1. Initial and final sound identification 2. Blending and segmenting words	1. Blending and segmenting nonwords 2. Making words

References

- Agocs, MM.; Burns, MS.; De Ley, LE.; Miller, SL.; Calhoun, BM. Fast ForWord Language.. In: McCauley, RJ.; Fey, ME., editors. Treatment of Language Disorders in Children. Paul H. Brookes; Baltimore, MD: 2006. p. 471-508.
- Aram DM, Hall NE. Longitudinal follow-up of children with preschool communication disorders: Treatment implications. *School Psychology Review* 1989;18(4):487–501.
- Beitchman JH, Wilson B, Brownlie EB, Walters H, Lancee W. Long-term consistency in speech/language profiles: I. Developmental and academic outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry* 1996;35:804–814. [PubMed: 8682762]
- Berger, V. Selection biases and covariate imbalances in randomized clinical trials. John Wiley & Sons; Chichester, England: 2005.
- Bernstein LE, Stark RE. Speech perception development in language-impaired children: A 4-year follow-up study. *Journal of Speech & Hearing Disorders* 1985;50(1):2130.
- Bishop DVM, Adams CV, Rosen S. Resistance of grammatical impairment to computerized comprehension training in children with specific and non-specific language impairments. *International Journal of Language & Communication Disorders* 2006;41:19–40. [PubMed: 16272001]
- Bishop DVM, Carlyon RP, Deeks JM, Bishop SJ. Auditory temporal processing impairment: Neither necessary nor sufficient for causing language impairment in children. *Journal of Speech, Language, and Hearing Research* 1999;42:1295–1310.
- Brinton B, Spackman MP, Fujiki M, Ricks J. What should Chris say? The ability of children with specific language impairment to recognize the need to dissemble emotions in social situations. *Journal of Speech, Language, and Hearing Research* 2007;50:798–811.
- Bryk, AS.; Raudenbush, SW. Hierarchical linear models: Applications and data analysis methods. 2nd Ed.. Sage Publications; London: 2002.
- Burlingame I, Sussman H, Gillam RB, Hay J. An investigation of speech perception in children with SLI on a continuum of formant transition duration. *Journal of Speech, Language, and Hearing Research* 2005;48:805–816.
- Camarata SM, Nelson KE, Camarata MN. Comparison of conversational-recasting and imitative procedures for training grammatical structures in children with specific language impairment. *Journal of Speech & Hearing Research* 1994;37(6):1414–1423. [PubMed: 7877298]

- Carrow-Woolfolk, E. Comprehensive Assessment of Spoken Language. American Guidance service; Circle Pines, MN: 1999.
- Catts HW, Fey ME, Tomblin JB, Zhang X. A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research* 2002;45:1142–1157.
- Cleave PL, Fey ME. Two approaches to the facilitation of grammar in children with language impairments: Rationale and description. *American Journal of Speech Language Pathology* 1997;6:22–32.
- Clegg J, Hollis C, Mawhood L, Rutter M. Developmental language disorders – a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry* 2005;46(2):128–149. [PubMed: 15679523]
- Cognitive Concepts. *Earobics Step 1* [Computer software].. Houghton Mifflin Company; Evanston, IL: 2000a. Cognitive Concepts. *Earobics Step 2* [Computer software].. Houghton Mifflin Company; Evanston, IL: 2000b.
- Cohen W, Hodson A, O'Hare A, Boyle J, Durrani T, McCartney E, Matthey M, Naftalin L, Watson J. Effects of computer-based intervention through acoustically modified speech (Fast ForWord) in severe mixed receptive-expressive language impairment: Outcomes from a randomized controlled trial. *Journal of Speech, Language, and Hearing Research* 2005;48:715–729.
- Conti-Ramsden G, Botting N. Classification of children with specific language impairment: Longitudinal considerations. *Journal of Speech, Language, Hearing Research* 1999;42(5):1195–1204.
- Conti-Ramsden G, Botting N. Social difficulties and victimization in children with SLI at 11 years of age. *Journal of Speech, Language, and Hearing Research* 2004;47:145–161.
- DiSimoni, F. The Token Test for Children. Riverside; Rolling Meadows, IL: 1978.
- Dorling Kindersley. *My First Amazing World Explorer* (Version 2.0) [Computer Software].. DK Interactive Learning; New York, NY: 1999.
- Ellis Weismer S, Evans JL. The role of processing limitations in early identification of specific language impairment. *Topics in Language Disorders* 2002;22(3):15–29.
- Fazio BB, Naremore RC, Connell PJ. Tracking children from poverty at risk for specific language impairment: A 3-year longitudinal study. *Journal of Speech & Hearing Research* 1996;39(3):611–624. [PubMed: 8783139]
- Fey ME, Cleave PL, Long SH, Hughes DL. Two approaches to the facilitation of grammar in children with language impairment: An experimental evaluation. *Journal of Speech & Hearing Research* 1993;36(1):141–157. [PubMed: 7680731]
- Gertner BL, Rice ML, Hadley PA. Influence of communicative competence on peer preferences in a preschool classroom. *Journal of Speech and Hearing Research* 1994;37:913–923. [PubMed: 7967575]
- Gillam RB, Crofford JA, Gale MA, Hoffman LM. Language change following computer-assisted language instruction with Fast ForWord or Laureate Learning Systems software. *American Journal of Speech-Language Pathology* 2001;10:231–247.
- Gillam RB, Loeb DF, Friel-Patti S. Looking back: A summary of five exploratory studies of Fast ForWord. *American Journal of Speech-Language Pathology* 2001;10:269–273.
- Gillam, RB.; Pearson, NA. Test of Narrative Language. Pro-ed.; Austin, TX: 2004.
- Gillam, RB.; Ukrainetz, TM. Language intervention through literature-based units.. In: Ukrainetz, TM., editor. Literate language intervention: Scaffolding PreK-12 literacy achievement. Thinking Publications; Eau Claire, WI: 2006. p. 59-94.
- Hartley DEH, Wright BA, Hogan SC, Moore DR. Age related improvements in auditory backward and simultaneous masking in 6- to 10-year-old children. *Journal of Speech, Language, and Hearing Research* 2000;43:1402–1415.
- Herbert JD, Gaudiano BA. Moving from empirically supported treatment lists to practice guidelines in psychotherapy: The role of the placebo concept. *Journal of Clinical Psychology* 2005;61:893–908. [PubMed: 15827997]
- Hill PR, Hogben JH, Bishop DVM. Auditory frequency discrimination in children with specific language impairment: A longitudinal study. *Journal of Speech-Language-Hearing Research* 2005;48:1136–1146.

- Hoffman LM, Gillam RB. Verbal and spatial information processing constraints in children with specific language impairment. *Journal of Speech-Language-Hearing Research* 2004;47:114–125.
- Johnston, JR. *Thinking About Child Language: Research to Practice*. Thinking Publications; Eau Claire, WI: 2006.
- Kaiser, AP.; Yoder, PJ.; Keetz, A. Evaluating milieu teaching.. In: Warren, SF.; Reichle, J., editors. *Causes and effects in communication and language intervention*. Brookes; Baltimore, MD: 1992. p. 948
- Kaufman, AS.; Kaufman, NL. *Kaufman Brief Intelligence Test*. American Guidance Service; Circle Pines, MN: 1990.
- Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *Journal of Clinical Epidemiology* 1999;52(12):19–26. [PubMed: 9973070]
- Knowledge Adventure, Inc.. *Dinosaur Adventure 3-D* (Version 4.0.1) [Computer Software].. Knowledge Adventure, Inc.; Torrance, CA: 1999.
- Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB. Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* 1996;273:971–973. [PubMed: 8688085]
- Kreft, I.; De Leeuw, J. *Introducing multilevel modeling*. Sage; London: 1998.
- Leonard, LB. *Children with specific language impairment*. MIT Press; Cambridge, MA: 1998.
- Littell, RC. *SAS for linear models*. 4th Ed. SAS Institute; Cary, NC: 2002.
- Lohr JM, DeMaio C, McGlynn FD. Specific and nonspecific treatment factors in the experimental analysis of behavioral treatment efficacy. *Behavior Modification* 2003;27:322–368. [PubMed: 12841588]
- Marler JA, Champlin CA. Sensory processing of backward-masking signals in children with language-learning impairment as assessed with the auditory brainstem response. *Journal of Speech, Language, and Hearing Research* 2005;48:189–203.
- Marler JA, Champlin CA, Gillam RB. Backward and simultaneous masking measured in language-learning impaired children who received intervention with Fast ForWord or Laureate Learning systems software. *American Journal of Speech-Language Pathology* 2001;10:258–268.
- Merzenich MM, Jenkins WM, Johnson P, Scheiner C, Miller SL, Tallal P. Temporal processing deficits of language-learning impaired children ameliorated by training. *Science* 1996;271:77–81. [PubMed: 8539603]
- Miller, J. *Systematic Analysis of Language Transcripts* (V. 7). Language Analysis Lab.; Madison, WI: 2003.
- Miller, L.; Gillam, R.; Peña, E. *Dynamic Assessment and Intervention of Children's Narratives*. Pro-Ed.; Austin, TX: 2001.
- Miller S, Linn N, Tallal P, Merzenich M, Jenkins W. Speech and language therapy (reeducation orthophonique). *Federation Nationale des Orthophonistes, Special Issue. La conscience phonologique* 1999;197:159–182.
- Nagarajan SS, Wang X, Merzenich MM, Schreiner CE, Johnston P, Jenkins WM, Miller S, Tallal P. Speech modification algorithms used for training language learning-impaired children. *IEEE Transactions on Rehabilitation Engineering* 1998;6:257–267. [PubMed: 9749903]
- Nelson KE, Camarata SM, Welsh J, Butkovsky L. Effects of imitative and conversational recasting treatment on the acquisition of grammar in children with specific language impairment and younger language-normal children. *Journal of Speech & Hearing Research* 1996;39(4):850–859. [PubMed: 8844564]
- Newcomer, PL.; Hammill, DD. *Test of Language Development:3-Primary*. Pro-Ed.; Austin, TX: 1997.
- Nordic Software, Inc.. *Coin Critters* (Version 2.1.4A) [Computer Software].. Nordic Software Inc.; Lincoln, NE: 1999.
- Norris JA. Providing language remediation in the classroom: An integrated language-to-reading intervention method. *Language, Speech & Hearing Services in the Schools* 1989;20:205–218.
- Norris, JA.; Hoffman, P. *Whole language intervention for school-age children*. Singular; San Diego, CA: 1993.

- Pokorni JL, Worthington CK, Jamison PJ. Phonological awareness intervention: Comparison of Fast ForWord, Earobics, and LiPS. *The Journal of Educational Research* 2004;97(3):147–157.
- Puelo JS, Pastore RE. Contralateral cuing effects in backward masking. *The Journal of the Acoustical Society of America* 1980;67:947–951. [PubMed: 7358919]
- Scholastic, Inc.. *The Magic School Bus Discovers Flight* [Computer Software].. Scholastic Inc.; New York, NY: 2001a.
- Scholastic, Inc.. *The Magic School Bus: Whales & Dolphins* [Computer Software].. Scholastic Inc.; New York, NY: 2001b.
- Scientific Learning Corporation. *Fast ForWord–Language* [Computer software].. Author; Berkeley, CA: 1998.
- Semel, E. *Following directions*. [Computer software].. Laureate Learning Systems; Winooski, VT: 2000.
- Semel, E.; Wiig, EH.; Secord, WA. *Clinical Evaluation of Language Fundamentals–*. 3rd Edition. The Psychological Corporation; UK. London, U.K.: 2000.
- Silva PA, Williams S, McGee R. A longitudinal study of children with developmental language delay at age three: Later intelligence, reading and behavior problems. *Developmental Medicine & Child Neurology* 1987;29(5):630–640. [PubMed: 2444484]
- Snowling M, Hulme C. A longitudinal case study of developmental phonological dyslexia. *Cognitive Neuropsychology* 1989;6(4):379–401.
- Snijders, TAB.; Bosker, RJ. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage Publications; London: 1999.
- Soleil Software, Inc.. *Zurk's Rainforest Lab* [Computer Software].. Soleil Software Inc.; Palo Alto, CA: 1998.
- St. Louis, KO.; Ruscello, DM. *Oral Speech Mechanism Screening Examination-Third Edition*. Pro-Ed.; Austin, TX: 2000.
- Stothard SE, Snowling MJ, Bishop DVM, Chipchase BB, Kaplan CA. Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech Language & Hearing Research* 1998;41(2):407–418.
- Strong, CJ.; Hoggan, KC. *The magic of stories: Literature-based language intervention*. Thinking Publications; Eau Claire, WI: 1996.
- Tallal P. Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language* 1980;9:182–198. [PubMed: 7363063]
- Tallal P. Fine-grained discrimination deficits in language-learning impaired children are specific neither to the auditory modality nor to speech perception. *Journal of Speech and Hearing Research* 1990;33(3):616–617. [PubMed: 2232779]
- Tallal P. Improving language and literacy is a matter of time. *Nature Reviews: Neuroscience* 2004;5:721–728.
- Tallal P, Gaab N. Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences* 2006;29:382–390. [PubMed: 16806512]
- Tallal P, Miller SL, Bedi G, Byrna G, Wang X, Nagarajan SS, Schreiner C, Jenkins WM, Merzenich MM. Language comprehension in language learning impaired children improved with acoustically modified speech. *Science* 1996;271:81–84. [PubMed: 8539604]
- Thibodeau L, Friel-Patti S, Britt. Psychoacoustic performance in Children completing Fast ForWord Training. *American Journal of Speech, Language Pathology* 2001;10:248–257.
- TLC Educational Properties LLC.. *Arthur's 1st Grade* [Computer Software].. The Learning Company; Novato, CA: 1999a.
- TLC Educational Properties LLC.. *Arthur's 2nd Grade* [Computer Software].. The Learning Company; Novato, CA: 1999b.
- Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith S, O'Brien M. Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research* 1997;40:1245–1260.
- Tomblin JB, Records NL, Zhang X. A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech & Hearing Research* 1996;39(6):1284–1294. [PubMed: 8959613]

- Tomblin JB, Zhang X, Buckwalter P, O'Brien M. The stability of primary language disorder: Four years after kindergarten diagnosis. *Journal of Speech, Language and Hearing Research* 2003;46:1283–1296.
- Wagner, RK.; Torgesen, JK.; Rashotte, CA. *Comprehensive Test of Phonological Processing*. Pro-Ed.; Austin, TX: 1999.
- Weismer SE. The role of stress in language processing and intervention. *Topics in Language Disorders* 1997;17(4):41–52.
- Wilson, MS.; Fox, BJ. *Micro-LADS* [Computer software].. Laureate Learning Systems; Winooski, VT: 1997.
- Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature* 1997;387(6629):176–178. [PubMed: 9144287]

Table 1
Pre-intervention Characteristics of the Children who Participated in Each of the Arms of the Study at Baseline

	Overall	CALI	Treatment Arms FFW-L	ILI	AE
Number of Participants	216	54	54	54	54
Age (years; months)	7;6	7;5	7;6	7;8	7;7
K-Bit Matrices	96.1 (9.1)	96.6 (9.0)	95.8 (9.2)	95.1 (9.5)	96.9 (8.7)
TOLD-P:3	73.8 (8.6)	75.6 (8.4)	73.0 (9.3)	73.9 (7.6)	72.6 (9.0)
Sex					
Male	136 (63%)	34	29	38	35
Female	80 (37%)	20	25	16	19
Race/Ethnicity					
White (Not Hispanic)	100 (46%)	25	31	24	20
Black/African-American	63 (29%)	15	12	17	19
White/Latino (Hispanic)	32 (15%)	9	4	10	9
Other	21 (10%)	5	7	3	6
Parental Education					
Completed college	81(37.5%)	23	20	19	19
Attended college	94(43.5%)	21	21	26	26
Did not attend college	41(19.0%)	10	13	9	9

Note. The standard deviations of the K-Bit Matrices and TOLD-P:3 measures are in parentheses. The TOLD-P:3 value is the Spoken Language Quotient.

Table 2
Average Number of Days at Each Compliance Level as a Function of Treatment Arm.

Compliance Level	Treatment Arm			
	FFW-L	CALI	ILI	AE
Level 1-verbal encouragement	3.06	4.61	6.04	2.3
Level 2-extra sticker	1.07	2.52	2.20	1.13
Level 3-short break	.89	2.0	.89	.96
Level 4-small food reward	.57	.43	.28	.50
Level 5-short time-out	.09	.04	.06	.09
Level 6-end intervention for the day	.02	.02	.02	.02
Level 7-meet with parents	0	0	0	0

Table 3

Number of cases (N), unadjusted means, and standard deviations (SD) for the CASL Composite standard scores and Backward Masking Thresholds as a Function of Treatment Arm and Time

Arm	Time	CASL: Composite Standard Score		Backward Masking (dB SPL)	
		N	Mean	SD	Mean
CALI	Pre-test	54	79.0	10.5	68.9
	Post-test	53	83.4	11.7	56.0
	3-month follow-up	52	84.9	12.8	54.1
	6-month follow-up	52	86.3	11.6	49.4
FFW-L	Pre-test	54	78.0	11.3	63.2
	Post-test	51	83.0	13.3	57.8
	3-month follow-up	51	83.8	13.7	51.6
	6-month follow-up	50	85.2	13.8	51.4
ILI	Pre-test	54	80.8	13.0	62.0
	Post-test	52	85.7	13.2	52.9
	3-month follow-up	51	87.8	13.2	51.8
	6-month follow-up	50	89.9	12.2	48.0
AE	Pre-test	54	78.2	11.1	62.6
	Post-test	53	82.1	11.4	55.2
	3-month follow-up	53	83.8	10.6	52.1
	6-month follow-up	53	86.2	10.8	49.3

Table 4

Unadjusted Means and Standard Deviations (SD) for CASL Subtest Raw Scores, Token Test Raw Scores, and Blending Words Raw Scores as a Function of Arm and Time.

Arm and Time	Antonyms		Syntactic Construction		Paragraph Comprehension		Subtests		Pragmatic Judgment		Token Test		Blending Words	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
<u>CALI</u>														
Pretest	13.5	5.5	17.9	6.7	24.2	11.4	19.1	9.5	21.3	8.3	8.0	4.5		
Posttest	14.7	5.2	19.5	7.1	28.6	11.7	21.9	10.2	22.3	8.2	9.0	3.9		
3 month	16.3	5.7	21.3	7.5	29.9	11.3	24.2	10.9	25.0	7.3	10.1	4.0		
6 month	17.1	4.7	21.7	7.3	33.6	8.1	26.1	12.21	25.6	7.7	10.6	3.9		
<u>FFW-L</u>														
Pretest	13.7	5.4	18.3	5.7	22.6	10.9	17.8	9.0	20.5	7.2	8.2	3.8		
Posttest	15.4	4.3	19.4	6.3	27.4	11.8	21.4	10.1	22.6	7.4	9.4	3.6		
3 month	16.3	4.5	20.5	6.4	30.9	8.2	22.7	11.4	25.3	7.4	9.9	3.8		
6 month	17.5	4.7	22.0	6.0	31.5	10.2	25.9	12.3	25.2	7.0	11.1	3.5		
<u>ILI</u>														
Pretest	15.4	6.3	19.7	5.4	22.0	15.5	21.2	10.0	22.7	6.7	8.7	3.6		
Posttest	16.6	4.9	21.3	5.3	28.4	13.4	23.5	11.1	23.5	7.3	9.6	3.1		
3 month	17.9	4.8	21.9	5.3	32.7	9.2	26.4	12.3	26.5	6.5	9.6	3.4		
6 month	18.7	4.4	24.3	5.4	34.5	7.7	30.4	11.8	26.6	7.3	10.4	3.1		
<u>AE</u>														
Pretest	14.1	6.1	17.9	6.5	22.9	12.3	18.8	8.1	21.4	7.5	8.3	3.8		
Posttest	15.8	5.2	19.6	6.1	28.3	12.3	20.4	9.5	21.6	8.8	7.9	3.6		
3 month	16.8	4.8	20.3	5.9	32.1	8.8	22.4	9.8	24.2	7.8	9.9	3.5		
6 month	17.9	4.5	22.3	6.1	34.4	8.3	24.8	9.9	24.1	7.9	9.9	3.1		

Cohen's d effect sizes from pretest to posttest and from pretest to 6-month follow-up for the CASL core composite, Backward Masking, the Token Test for Children, the blending words subtest on the C-TOPP.

Table 5

Arm	CASL		Backward Masking		Token Test		Blending words	
	Posttest	6 month	Posttest	6 month	Posttest	6 month	Posttest	6 month
CALI	.56	1.02	.71	1.14	.12	.54	.26	.62
FEW-L	.71	.93	.29	.65	.29	.66	.32	.79
ILI	.79	1.33	.51	.79	.11	.56	.27	.53
AE	.61	1.34	.39	.76	.02	.35	-.11	.46

Table 6

Percentage of Children Whose CASL Composite Scores Were Outside the 68% (1 SEM) and 95% (1.96 SEM) Confidence Intervals.

Arm	68% Confidence Interval		95% Confidence Interval	
	Immediate Posttest	6-Month Follow-up	Immediate Posttest	6-Month Follow-up
CALI	55.6%	63.0%	31%	54%
FFW-L	53.7%	74.1%	33%	59%
ILI	59.3%	79.6%	43%	69%
AE	53.7%	68.5%	31%	54%