A Preliminary Evaluation of Fast ForWord-Language as an Adjuvant Treatment in Language Intervention

Marc E. Fey
Lizbeth H. Finestack
Byron J. Gajewski
Mihai Popescu
University of Kansas Medical Center, Kansas City

Jeffrey D. Lewine
Alexian Brothers Medical Center, Elk Grove Village, IL

Purpose: Fast ForWord-Language (FFW-L) is designed to enhance children's processing of auditory–verbal signals and, thus, their ability to learn language. As a preliminary evaluation of this claim, we examined the effects of a 5-week course of FFW-L as an adjuvant treatment with a subsequent 5-week conventional narrative-based language intervention (NBLI) that targeted narrative comprehension and production and grammatical output.
Method: Twenty-three children 6–8 years of age with language impairments were assigned randomly to 1 of 3 intervention sequences: (a) FFW-L/NBLI, (b) NBLI/FFW-L, or (c) wait/NBLI. We predicted that after both treatment periods, the FFW-L/NBLI group would show greater gains on measures of narrative ability, conversational grammar, and nonword repetition than the other groups.
Results: After the first 5-week study period, the intervention groups, taken together (i.e., FFW-L/NBLI and NBLI/FFW-L), significantly outperformed the no-treatment wait/NBLI group on 2 narrative measures. At the final test period, all 3 groups displayed significant time-related effects on measures of narrative ability, but there were no statistically significant between-groups effects of intervention sequence.
Conclusions: This preliminary study provides no evidence to support the claim that FFW-L enhances children's response to a conventional language intervention.

KEY WORDS: Fast ForWord, language intervention, child language disorders, narratives

There is evidence supporting the efficacy of interventions that teach new vocabulary, grammatical forms, and narrative content and structure to children with language impairments, especially when the children's problems are in expression rather than comprehension (Cirrin & Gillam, 2008; Fey, Cleave, Long, & Hughes, 1993; Gillam, McFadden, & van Kleeck, 1995; Law, Garrett, & Nye, 2004; Leonard, Camarata, Brown, & Camarata, 2004; K. E. Nelson, Camarata, Welsh, Butkovsky, & Camarata, 1996). For the most part, the basic strategy underlying these successful interventions is to treat the symptoms of the disorder through modification of the language-learning environment. The true promise of developmental language intervention, however, lies not in its ability to teach language to children in a piecemeal fashion but in its potential to prepare children to learn language more efficiently on their own from ambient sources. At present, there is no evidence that conventional, compensatory forms of language intervention boost the child's ability to learn outside the teaching context. To the contrary, the limited evidence available suggests that children with language impairment continue to exhibit deficits in oral and written language throughout the school years even when they receive language intervention as part of

their preschool and elementary educations (Fey, Catts, & Larrivee, 1995; Tomblin, Zhang, Buckwalter, & O'Brien, 2003).

Rather than directly teaching language form, content, and use, Fast ForWord-Language (FFW-L) is a different type of language intervention that targets what its developers assume is the very basis of children's language impairments: an inability to properly process the rapidly changing acoustic features of the speech stream. The program includes seven computerized listening games with acoustically modified nonspeech and speech stimuli. Consistent with its goals, the key variables manipulated in the program are acoustic rather than linguistic. Tallal (2000) explained that "the goal was to drive, through adaptive training, each child into the normal processing rate of tens of milliseconds, while simultaneously increasing each child's ability to process linguistic structures in their most frequent, naturally occurring, obligatory contexts" (p. 145). Consequently, what is affected should be the child's ability to learn language and to learn from language. The developers and marketers of FFW-L, the Scientific Learning Corporation (SLC), stated this succinctly in their description of the latest version of the program on their Web site: "And, most importantly, the gains students achieve are lasting, the result of enduring positive changes in their processing skills and learning capacity" (SLC, 2009, para. 3). The present investigation is an exploratory effort to examine this potential effect.

Previous Evaluations of FFW-L

FFW-L has undergone a number of nonexperimental and experimental evaluations as a primary intervention with the expectation that participation would lead directly to better performance on language tasks. Early quasi-experimental studies implemented by the developers of the program reported dramatic improvements in speed of auditory processing assumed by the investigators to play a key role in developmental language impairment. In addition, large gains were observed in performance on standardized language tests over intervals of intensive training as short as 4 weeks (Merzenich et al., 1996; Tallal, 2000; Tallal et al., 1996). As encouraging as they were, however, these gains in language were not interpreted by Tallal et al. (1996) as the product of enhanced learning ability. Tallal et al. astutely observed the following:

> It seems unlikely that these children learned the equivalent of approximately 2 years of language in 1 month. Rather, it appears that they had already developed considerably more language competence than they were able to demonstrate or use "on line" under normal listening and speaking conditions. (p. 83)

In this study, then, the FFW-like games that the children played may have made it easier for them to utilize existing language skills. The impact of these games on learning efficiency or potential, however, was not tested.

More recent randomized trials of FFW-L have reported either no treatment-related effects on language test performance (Pokorni, Worthington, & Jamison, 2004; Rouse & Krueger, 2004) or clinically significant gains shown not to be specific to FFW-L (Cohen et al., 2005; Gillam et al., 2008). Because FFW-L games require listening only, with no production exercises, it may seem better suited for children with receptive language problems or for children with confirmed problems with auditory processing. Studies involving children with comprehension or auditory backward masking problems (Cohen et al., 2005; Gillam et al., 2008; Thibodeau, Friel-Patti, & Britt, 2001), however, have not shown FFW-L to be selectively efficacious for these groups.

As for earlier studies, these more recent investigations generally have been designed to measure the direct effects of FFW-L as an isolated treatment on language performance rather than as an agent designed to enhance or supplement the effects of other interventions. Consequently, they have placed no controls on the language input children have received after completion of the program.

One recent study (Stevens, Fanning, Coch, Sanders, & Neville, 2008) offers one kind of exception to this general trend by measuring a physiologic response to FFW-L. Stevens et al. (2008) observed significant gains on a standardized language comprehension test among children who had participated in 6 weeks of FFW-L. More importantly, however, they noted an effect on the brain activity associated with selective attention, as measured using event-related brain potentials. It is possible that improved selective auditory attention sharpens children's language-learning ability, thus leading to greater learning efficiency following FFW-L. Alternatively, improved attentional skills could result in better standardized test performance with limited or no effects on learning potential. However, Stevens et al.'s study was not designed to distinguish between these two possibilities.

In sum, the effects of FFW-L as an isolated language intervention have been evaluated in an impressive number of studies—the most recent and well controlled of which have observed either limited treatment effects on language performance or significant pre–post improvements that are not specific to FFW-L. Although some of these studies have directly examined the effects of FFW-L on attentional and auditory processes that may be crucial to language learning (Gillam et al., 2008; Merzenich et al., 1996; Stevens et al., 2008; Tallal et al., 1996), none of these studies have directly addressed the most central

claim of the SLC—that FFW-L makes children better language learners. A systematic assessment of such an effect requires control over the input that the children receive subsequent to participation in FFW-L so that the products of posttreatment learning can be examined and compared with that of children who have not received FFW-L. One way of exercising such control is to test FFW-L's effects as an adjuvant treatment designed to boost children's response to a conventional language intervention. If FFW-L changes children's processing skills and learning capacity, it should enhance their response to an intervention sequenced after its administration. This premise forms the underlying motivation for the primary question and design of the present study.

Questions and Predictions of the Current Study

For the present study, we posed two basic questions based on the language performance of 6- to 8-year-old children with language impairments. First, does a 5-week course of FFW-L lead to immediate gains in narrative and grammatical abilities that are greater than the gains associated with a low-intensity dose of a conventional narrative/grammar intervention or a no-treatment control period? Second, and more importantly, are outcomes associated with a low-intensity narrative and grammatical intervention boosted significantly when the intervention follows FFW-L?

To address these questions, we modified a conventional intervention—narrative-based language intervention (NBLI; Finestack, Fey, Sokol, Ambrose, & Swanson, 2006; Swanson, Fey, Mills, & Hood, 2005)—which had previously been tested as an 18-session, 6-week, individual treatment, so that it could be completed in 12 sessions over a 5-week period in a small group format. We compared the effects of this intervention across three intervention sequences: (a) the idealized sequence of a 5-week course of FFW-L followed by NBLI (FFW-L/NBLI), (b) a nonidealized sequence of NBLI followed by FFW-L (NBLI/FFW-L), and (c) a 5-week wait period without treatment followed by NBLI (wait/NBLI). Our primary intention was to measure the effects of FFW-L on the outcomes of our modified NBLI in two areas targeted directly by NBLI (comprehension and production of narrative language and the use of complex grammar) and one area of language processing that may be addressed more directly by FFW-L (phonological working memory). We administered our outcome measures preexperimentally (i.e., Time 1) after the first 5-week experimental period (i.e., Time 2) and again after the second experimental period (i.e., Time 3). The entire sequence was delivered over the summer to rule out effects of background interventions.

On the basis of the results of the most rigorous clinical trials of FFW-L to date, we predicted that at Time 2, any advantages of the FFW-L/NBLI group over the other groups would be limited. We expected percentage phonemes correctly repeated on the Nonword Repetition Test (NRT; Dollaghan & Campbell, 1998) to be perhaps the most likely measure to reveal an FFW-L advantage at this time because the NRT challenges memory processes that appear to be targeted directly by several FFW-L games.

More importantly, however, we predicted that at Time 3, the children who received the FFW-L/NBLI sequence would significantly outperform the NBLI/FFW-L and wait/NBLI groups on measures of narration, grammar, and nonword repetition. To meet the specifications of the grant funding the project, we needed to complete all interventions and testing over two summers at a single site where interventions could be monitored carefully. Thus, we had significant limitations in the number of children who could be involved, and we anticipated that power for our primary analyses would be low. Consequently, we regarded the study as exploratory and proposed from the outset to use any encouraging findings with respect to our hypotheses and predictions as the basis for developing a larger, fully powered randomized clinical trial.

Method

Participants

Recruitment

Children were recruited in two waves between January and May of 2006 and 2007, so that all interventions could be provided over each year's respective summer with no overlap with the school year. The original plan for recruiting participants was to include only children referred by speech-language pathologists in the Kansas City, KA, public schools (U.S. District 500). This would have made it possible to provide transportation to the clinical site at the University of Kansas Medical Center (KUMC), which also is in Kansas City, for all participants. U.S. District 500 is a large, urban district with a high concentration of minority students. According to the district, 44% of its students are African American, 35% are Hispanic, 17% are Caucasian, and 4% represent other ethnic groups. Of the students, 76% qualify for free or reduced lunch. Despite the cooperation and assistance of the U.S. District 500 schools, additional recruiting from the broader Kansas City metro area was necessary in 2007 to attain the numbers of children needed and to obtain a larger number of children who were Caucasian and not Spanish speaking.

Referral Criteria

Referrals were requested for children finishing kindergarten, first grade, or second grade who either were

on existing caseloads for speech and language or reading services or who had been identified as being at risk and were being closely monitored by a special education team to evaluate the need for speech and language services. We asked for referrals of children who had not been diagnosed with intellectual disability. Only children with normal hearing and no history of neurological disorder—such as stroke, closed head injury, or seizure disorder—were included. English was required to be the primary language in the children's homes, thus eliminating a large proportion of Hispanic children in U.S. District 500.

All children who were referred were first tested in their homes, using a standardized qualifying test battery. This battery included the Test of Language Development–Primary, Third Edition (TOLD–P:3; Newcomer & Hammill, 1997), the Goldman Fristoe Test of Articulation—Second Edition (GFTA–2; Goldman & Fristoe, 2000), and the Matrices subtest of the Kaufman Brief Intelligence Test, Second Edition (KBIT–2; Kaufman & Kaufman, 2004). To qualify, children had to have Spoken Language Quotients on the TOLD–P:3 of 80 or lower (*SD*s = –1.33) and TOLD–P:3 and Listening Quotients at least 5 points (roughly 1 *SE*) below the mean of 100 (i.e., 95) to increase the likelihood that children's true scores were below average in language comprehension. They also had to have GFTA–2 scores of 81 or above and KBIT–2 nonverbal standard scores above 70.

Randomization of Participants and Blinding of Study Personnel

In total, 60 children were referred. After they assented and their parents signed consents to participate that were approved by the KUMC Human Subjects Committee, they were tested for the summer program, which we called "KUMC Story Camp." Thirty children qualified and were assigned to one of the three treatment groups: (a) FFW-L/NBLI, (b) NBLI/FFW-L, or (c) wait/NBLI. At the beginning of each summer study period, a research associate not affiliated with this project randomly determined group placements using a computerized tool developed for this purpose. The treatment protocols and facilities limited the number of participants who could be enrolled in each of the treatment groups to five participants. Therefore, participants were randomly assigned to one of the three treatment groups until one of the treatment groups was filled. The remaining participants were then assigned to one of the two other groups using the computerized assigner until filled. Three sibling pairs qualified for the study. Randomization was restricted so both members of each pair were assigned to the same group. The randomization process placed one pair in each group.

To complete adequate transportation and treatment plans on time, it was necessary to inform the investigators and interventionists of the participants' group assignments before preexperimental testing began. On test days, these personnel were available to transport children from one test session to another at KUMC, to keep children engaged as they waited for their next test, and to prepare and serve lunch; however, they did not participate in testing the children. Testers, test scorers, transcribers, language sample analyzers, and reliability checkers were kept blind to group assignments throughout the entire experimental and analysis periods.

Preexperimental Group Comparisons

Of the 30 qualifying children assigned to treatment groups, only 25 attended 50% or more of the sessions for at least one of the assigned interventions, and only 23 attended 50% or more of the scheduled sessions for both interventions. Because this study was planned to be exploratory, and we knew it was underpowered, it was not appropriate to rob additional power from it by performing an intention-to-treat analysis, which would include all identified participants in all analyses. Thus, the analyses presented in this report focus on only the 23 children who attended at least half of their planned sessions.

The participants' performance on criterion variables, their preexperimental scores on two primary dependent measures, and key demographics are provided in Table 1 . All participants had notable deficits in expressive language, as illustrated by their scores on the Speaking Quotient of the TOLD–P:3. Their impairments were generally more severe in expression than in comprehension, and their nonverbal IQs were generally somewhat below average, as is common among groups of children ascertained for primary language impairment (Stark & Tallal, 1981). Thus, on the basis of their borderline nonverbal IQ scores, 11 of these 23 children had nonspecific language impairment, and 12 had specific language impairment (Catts, Fey, Tomblin, & Zhang, 2002).

Key demographic data also are provided in Table 1. Consistent with the characteristics of the school district, most qualifying children were African American. In fact, all children in the FFW-L/NBLI group were African American. Most children had mothers who completed Grade 12, but few had more than 1 year of postsecondary education. Most of the children had been receiving language and/or reading services at school, and approximately half of the children had family members also being treated for speech and language difficulties.

In general, comparisons of the preexperimental data across groups indicate that our randomization successfully divided the sample into subgroups that did not differ

Table 1. Preexperimental means and standard deviations for key participant variables.

Variable	FFW-L/NBLI (*n* = 7)			NBLI/FFW-L (*n* = 7)			Wait/NBLI (*n* = 9)			FFW-L/NBLI: WD (*n* = 4)			NBLI/FFW-L: WD (*n* = 3)		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Age (years)	7.41	0.51		7.76	0.67		7.57	0.61		7.27	0.38		7.64	0.34	
TOLD–P:3 Speaking Quotient (SS)	69.14	9.44		65.71	14.16		71.67	11.73		73.00	10.95		70.00	15.00	
TOLD–P:3 Listening Quotient (SS)	77.71	9.14		75.57	7.83		83.67	8.50		79.00	4.24		84.00	6.25	
KBIT–2 nonverbal IQ	81.14	4.71		83.29	9.18		89.11	12.30		80.50	4.20		78.00	6.25	
GFTA–2 (SS)	100.00	5.88		96.43	4.54		99.54	5.15		91.00	10.23		102.00	3.61	
TNL NLAI (SS)	73.86	13.72		70.86	9.12		74.33	8.63		70.75	7.09		73.00	16.70	
NRT	68.15	7.51		56.70	17.64		63.43	7.11		64.32	12.10		53.12	18.43	
Maternal education (years completed)	12.86	0.69		13.14	3.08		13.33	0.87		12.50	0.58		9.67	1.16	
African American			7			5			5			3			2
Caucasian			0			2			4			1			1
In language-related services (child)			5			6			7			4			3
Record of language-related services (family)			2			4			6			3			3

Note. FFW-L = Fast ForWord-Language; NBLI = narrative-based language intervention; WD = children who were tested and assigned to groups but withdrew and were lost to follow-up; TOLD–P:3 = Test of Language Development—Primary, Third Edition; SS = standard score; KBIT–2 = Kaufman Brief Intelligence Test, Second Edition; GFTA–2 = Goldman Fristoe Test of Articulation—Second Edition; TNL = Test of Narrative Language; NLAI = Narrative Language Ability Index; NRT = Nonword Repetition Test.

substantially at Time 1. One-way analyses of variance (ANOVAs) for continuous measures and chi-square analyses for nominal variables identified no statistically reliable differences between the groups ($ps > .10$). Although the difference between groups was not statistically significant, the NBLI/FFW-L group mean was substantially lower than the other subgroups' means on the NRT. This group had one outlier. When this child was removed from the analysis, the mean for this group was 62 ($SD = 11.73$), bringing it well into line with the other groups. This child was excluded from all analyses using NRT as a dependent measure.

The two right-most columns in Table 1 include the preexperimental data for the seven children who were excluded from all analyses. All seven children were assigned originally to receive both treatments. Data for five of these seven children were lost because of failed follow-up after repeated efforts to phone and visit the residences of their caregivers. The other two children participated in more than half of the first of their two assigned interventions but not in the second. An examination of Table 1 indicates that these children who dropped out of the study were not obviously different from each other or the children in the larger subgroup to which they had been assigned.

Experimental Testing and Dependent Measures

Each summer of the study, all study participants came to KUMC three times for a full day on which all Time 1, Time 2, and Time 3 testing was completed. Testers were unfamiliar with the children at Time 1. At each test point, the testers collected a 20–30-min language sample using the conversational narrative elicitation procedures of Leadholm and Miller (1992), and they administered the Test of Narrative Language (TNL; Gillam & Pearson, 2004) and the NRT (Dollaghan & Campbell, 1998). Across all participants and at each test time, the order of tests was kept constant (i.e., language sample, followed by the TNL, followed by the NRT). Participants were given breaks between tests. At each testing time, the same examiners administered all of the assessments to their assigned participants. No children had the same examiner across two consecutive test times, and examiners had no additional contacts with any of the participants. On each testing day, children also were taken to the Hoglund Brain Imaging Center at KUMC for neurophysiologic testing. The results of the EEG outcomes are presented in another report (Popescu, Fey, Lewine, & Popescu, 2009).

The primary outcome measure for the study was the Narrative Language Ability Index (NLAI) from the TNL. The TNL is a standardized instrument for evaluating the story comprehension, retell, and formulation skills of children 5–11 years of age. Six subtests require the child to make up, retell, and answer questions about stories. These skills are taught in an explicit manner in the NBLI intervention. The NLAI is a standard score that reflects a composite of three narrative comprehension and three oral narration subtests. Therefore, our strongest prediction was that at Time 3, the FFW-L/NBLI

group would have a significant advantage over the other two groups for this measure or, given the auditory focus of FFW-L, at least for its narrative comprehension subcomponent.

In studies to date, narrative interventions have had inconsistent effects on the narrative performance of school-age children. For example, Cable (2007) observed large positive effects on oral narration of second graders with weak narrative skills, as measured by the Oral Narration composite of the TNL. In contrast, her narrative intervention had no effects on narrative comprehension. Westerveld and Gillon (2008), whose treatment program taught narrative structure to children 7–9 years of age with reading disabilities, observed precisely the opposite pattern. Their treatment yielded effects for narrative comprehension but not production.

To help us identify any such asymmetrical patterns underlying observed effects for NLAI, we planned to perform analyses on the raw scores for both the Narrative Comprehension (Comp-raw) and the Oral Narration (Oral-raw) composites of the TNL. We expected these analyses to be revealing because they not only examine comprehension and production separately but the raw scores do not reflect a comparison with the normative sample. Thus, they are a more direct indication of each group's change in narrative performance across time.

Two measures for grammar based on the language sample data were included as dependent measures: the mean main verb score per sentence from the Developmental Sentence Score (DSS; Lee, 1974) and the proportion of sentences containing a conjunction other than "and" or "then." Increases in DSS main verb scores occur when children use more complex and later-developing forms of verbs, such as auxiliary constructions and verb inflections in main clauses, and when they produce sentences with more than one main clause. Both standard DSS and Black English Sentence Scoring (BESS; N. W. Nelson & Hyter, 1990) were used for all children judged to be using African American English (AAE). BESS gives children credit for use of AAE forms that would otherwise be penalized in standard DSS (e.g., omission of copula and auxiliary "be"), and it yields substantially higher scores than DSS for AAE users, especially in the main verb category. For our purposes, however, the lower main verb scores from the standard DSS offered greater potential for change across time for our AAE speakers. Therefore, we ultimately used DSS rather than BESS for our analyses of verb usage for all participants.

Grammatical constructions involving conjunctions other than "and" and "then" include coordinators (i.e., "but," "or") and subordinators (e.g., "before," "after," "when," "since," "because"). These forms are needed to reflect increasingly complex relationships involving characters across time in stories, and they are known to be present in the stories of African American children and to be teachable to preschool-age African American children (McGregor, 2000). These measures were selected because compound and complex sentences containing more than one clause, the use of auxiliary verbs, and the use of coordinators and adverbial subordinators were the most common grammatical targets for the children's NBLI interventions. We assumed that these measures would be equally sensitive to changes made by participants who were African American and Caucasian. Consequently, we predicted that by Time 3, the FFW-L/NBLI group would exhibit greater improvements on these measures of grammatical production.

We were less confident of our predictions for the grammatical measures, however, than that for NLAI. Fey et al. (1993) observed treatment effects on DSS-based measures of grammar in conversation following a 20-week grammar intervention provided to preschoolers with grammatical impairments. In contrast, the 6-week NBLI treatment that Swanson et al. (2005) provided to 7- and 8-year-old children with language impairment did not lead to gains in grammatical production in conversation, as measured by the DSS. Loeb, Stoke, and Fey (2001) provided home-based FFW-L to four children with language impairments, but only one demonstrated significant gains on the DSS. Thus, the sensitivity of the DSS and other measures of spontaneous grammatical output to short-term interventions has not been well established.

The total percentage of phonemes correct (PPC) from the NRT (Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000) was calculated at each testing as an additional outcome measure. The NRT requires children to listen to and repeat 16 nonce words varying in length from one to four syllables. This type of test is frequently regarded as a measure of phonological working memory, which may have a significant role in word learning (Baddeley, Gathercole, & Papagno, 1998). Because FFW-L contains games devoted to enhance memory, we regarded changes in PPC as a relatively direct outcome of FFW-L, especially as compared with the narrative and grammatical language measures. When used with other indicators of developmental language impairment, PPC can be a useful, culturally unbiased marker of language impairment in samples, including African American children (Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Oetting & Cleveland, 2006). On the NRT, children judged to be making frequent use of grammatical and phonological features of AAE were given credit for devoiced final consonants and for nondiphthongized versions of the vowel diphthongs (Oetting & Cleveland, 2006), even though the speech of our participants was generally good on a citation form test (GFTA–2), and such corrections were found to be unnecessary by Dollaghan and Campbell (1998) and by Oetting and Cleveland (2006). Thus, these

scores may slightly overestimate the children's performances on this test.

Transcription of Language Samples

The language sample and Oral Narration subtests from the TNL were recorded using Marantz PMD660 digital tape recorders and an Audio-Technica 8531 microphone system. Samples were transcribed by graduate assistants who were pursuing their master's degrees in speech-language pathology, with Sennheiser HD 280 Pro headphones, using Systematic Analysis of Language Transcripts software (Miller & Chapman, 2000) conventions. Two assistants were assigned to each transcript. As part of a consensus transcription procedure, each assistant first transcribed each assigned sample independently of the other transcriber. Prior to consultation with the other transcriber, one assistant compared the two independent transcriptions, marking any disagreements between judges on words, morphemes, and utterance boundaries in the samples. Assistants then made changes to the transcripts in an effort to reach a consensus. Words and morphemes on which agreement could not be reached were marked as unintelligible.

A set of 26 (~30% of the total) of these samples was selected at random to calculate transcriber agreement prior to consultation between partners. Expressed as the ratio of agreements to agreements plus disagreements, agreement was 96% (range = 90%–97%) for words, 92% (range = 78%–98%) for bound morphemes, and 93% (range = 82%–100%) for utterance segmentation.

Reliability of Measures

Scoring of the primary outcome tests—the TNL, DSS, and NRT—was carried out by graduate assistants in speech-language pathology or American Speech-Language-Hearing Association (ASHA)-certified speech-language pathologists, all of whom were blind to the children's group assignments. Approximately 60% of all TNL samples, 20% of the DSS samples, and all of the NRT data were scored by two judges working independently. Reliability for the dependent measures was estimated using intraclass correlation coefficients (ICCs). Because these tests were not used as absolute diagnostic criteria for our study, the consistency definition of the ICC was used. This measure involves a Judges × Participants ANOVA and reflects a calculation of the proportion of variance in each dependent variable at the time of measurement that can be attributed to systematic differences between participants (i.e., true participant differences) and to interactions between judges and participants. ICC values can range from 0 to 1.0; those closer to 1.0 indicate that a greater proportion of the variance is associated with systematic participant differences with less of an impact of judges on participant scores. The ICCs ranged from .87 to .99. These very high values for all measures indicate that most of the variance in the outcome measures was due to the systematic differences between the participants at the time of measurement, with relatively small amounts attributable to the impact of judges on the participant scores.

Interventions

FFW-L

Comprehensive descriptions of the FFW-L computer games are available elsewhere (Agocs, Burns, DeLey, Miller, & Calhoun, 2006), so they are not described in detail here. Briefly, there are seven FFW-L exercises, each embodied as a computer game designed to maintain participant interest and to challenge the child's auditory processing skills. The key to each game is its adaptive nature. Over the course of game play, nonspeech stimuli are increased in complexity by shortening their overall durations and interstimulus intervals. Initially, speech stimuli are presented at reduced rates with intensified formant transitions. With successful performance, the rates are increased, and the intensity of formant transitions is decreased, thus rendering the speech stimuli more like natural speech. Each game is embedded within a graphics interface that resembles a traditional video game, and children score points for correct performance. Three of the seven games focus on basic aspects of auditory processing (*Circus Sequence, Old MacDonald's Flying Farm,* and *Phoneme Identification*), two have at least some lexical component (*Phonic Match* and *Phonic Words*), and two are clearly verbal in nature (*Language Comprehension Builder* and *Block Commander*).

The children assigned to receive FFW-L were scheduled to attend 24 daily sessions for a period of approximately 5 weeks. When children arrived for therapy, they started immediately playing one of the five games that the program selected for them on that day. They played three games, took a snack break, and then continued to play the remaining two games. Children played at their own computers in a room with no more than five children for a total of 100 min for each scheduled session. Although we made efforts to make the computer games rewarding and fun, session length was tightly controlled, so the time for having fun and interacting informally with the adults and other children was limited.

All FFW-L sessions were monitored by the first and second authors, both of whom are certified speech-language pathologists and were trained by the SLC to implement FFW-L. The main function of this monitoring was to prepare individualized performance charts, to praise children for reaching their point goals, and to keep the children on task while working toward their goals. Poker chips were awarded for reaching daily performance

targets and later exchanged for small prizes at the end of each day. Unlike the procedures in some studies (e.g., Gillam et al., 2008), when children had special difficulty on a particular game, we gave specific feedback and direct teaching to help children to understand the keys to playing the games, as was recommended by the SLC. For example, when playing the most challenging game, *Circus Sequence,* children frequently seemed not to understand the basic goal of the game, especially when the program removed visual cues on which many children relied. As recommended by SLC staff, we instructed these children to close their eyes and otherwise complete the early stages of the game without the visual cues. If this or other online cuing did not result in better performance, children were taken offline for short periods to practice playing discrimination games using live voice models or a xylophone.

NBLI

NBLI was designed to be a relatively brief (i.e., 5-week) complement to FFW-L that would give children with language impairment idealized exposure to, production experience with, and direct instruction in narration and grammar—aspects of language known to cause them difficulty (Finestack et al., 2006; Swanson et al., 2005). Children attended two NBLI sessions one week and three sessions the next week in alternating fashion over the 5-week period, for a total of 12 planned sessions for each child. The basic narrative treatment goal for all participants was to be able to develop and orally produce a coherent story containing basic story elements, such as the beginning (i.e., character and setting development), a clear problem or goal, actions taken to resolve the problem and their consequences, and a story resolution. Grammatical goals included coordinate and subordinate clauses and conjunctions, noun postmodification (e.g., relative clauses, appositives), elaborated verbs (i.e., verb phrases containing auxiliaries), and—for children clearly not speaking AAE—regular past tense. Over the 5-week intervention course, two grammatical goals were selected for each participant. These goals were alternated on a weekly basis.

Fictional stories and accompanying pictures, developed for this and related projects, were central to each NBLI session. During each session, one of 18 available stories was selected and presented to the treatment group. Each story provided a clearly delineated model of the targeted narrative structure as well as a rich supply of grammatical features commonly selected as goals. Stories and activities were developed so that each story could serve as a model for several different grammatical goals. Each NBLI session had four basic activities: warm-up, story retell, sentence imitation, and story generation.

Warm-up activity. During this initial activity, participants were given the opportunity to retell a story with little or no correction by the clinician. The children were familiar with the stories used in this task from the story retell–imitation task (see below) from the previous session. The activity was designed to get each session off to a successful start and to provide the children the opportunity to incorporate key story components representing the beginning (i.e., characters and setting), problem/goal (i.e., initiating event, reaction, plans), actions (i.e., attempts and consequences), and ending in their retell. In our group delivery of the intervention, the children took turns and were encouraged to support one another in this retell activity.

Story retell. In this task, the children retold a new story that contained multiple examples of the grammatical and narrative goals for each child. The clinician first read the story in its entirety. Then, she read it again, one story component at a time. Each child was asked to paraphrase a part of the story immediately after the clinician read it. If the child produced a grammatical error on a target form or a sentence that could have included a target form to good effect, the clinician recast the child's sentence by correcting the error but did not ask for a repetition. The clinician was required to produce at least five recasts specific to each child's daily goal per 60-min session, although the ultimate target was 10 recasts. The clinician's primary goal, however, was to recap the children's story contributions to ensure that they included all crucial story components in the appropriate order. Thus, the structure and content of the story were of principal importance.

Sentence imitation task. A short sentence imitation drill was designed to provide the children with intensive exposure to and production practice with his or her grammatical targets and to force the child to attend closely to the grammatical details of clinician models. The 10–12 target sentences were taken directly from the story the children retold in the retell task. Thus, some textual detail was present even in this imitative exercise. The clinician was mandated to get each child to produce at least four target sentence imitations per session, although many more were typical.

Story generation task. For this final task, the clinician showed the children a picture representing many actions within a common theme (e.g., at the beach) without an obvious initiating event. The clinician and children generated and discussed a problem or potential problem represented in the picture and then used it as the basis for the story. The children and clinician then cocreated a story, including the story elements established and practiced in the story retell and warm-up activities. As the children developed new story components, the clinician queried them for crucial information

that they may have omitted. She also recapped story details, thus highlighting crucial story components, such as the setting, the problem, attempts to resolve the problem, the consequence, and the ending. The clinician captured story details in stick drawings, which were then used to help the children tell the story again on their own (Ukrainetz, 1998).

NBLI session schedule. Our research questions did not require us to match the level of intensity for FFW-L with our NBLI sessions. However, most children in each group came to their scheduled sessions and returned home on the bus. Because NBLI sessions were scheduled for only 60 min, we had to provide extra activities to occupy NBLI participants for an additional 60 min until the bus arrived. Therefore, we developed a set of standard activities to precede and follow the bona fide NBLI sessions, including a snack, listening to stories, and drawing and coloring. These activities were entertaining and fun, and although some might well have contributed to any effects of NBLI, they were not designed to do so. The clinician was instructed never to use NBLI teaching techniques—such as requests for sentence imitation, sentence recasts, requests for story retelling, discussion of story parts, and so forth—in beginning and coda activities.

Treatment Exposure and Fidelity

For FFW-L, the games to be played, the time played, the stimuli and their acoustic modification, and the contingencies for moving to higher levels of the program are built into the program. Therefore, fidelity of intervention with FFW-L is reflected in large part in the figures for attendance and the number of minutes completed. These values, which are recorded and maintained by the FFW-L program, are presented for each treatment group as well as for individual participants in Table 2. On average, of the 24 sessions offered, participants in the FFW-L/NBLI and NBLI/FFW-L groups attended 20 and 22 sessions, completing approximately 1,900 and 2,200 min of treatment, respectively. These values do not differ significantly across groups, $ts(12) = -1.38$, $p = .19$.

Table 2. Records of attendance, FFW-L program completion, and NBLI child attention ratings.

Group	Child	No. of FFW-L sessions attended	Total FFW-L minutes	No. of NBLI sessions attended	NBLI clinician	Mean NBLI attention rating	% of FFW-L completed
FFW-L/NBLI							
	1	15	1,485	6	2	2.96	67
	2	23	2,273	7	1	2.33	35
	3	23	2,254	6	1	2.33	12
	4	19	1,846	10	4	2.86	72
	5	23	2,246	12	4	2.86	53
	6	22	2,040	11.5	3	2.88	36
	7	15	1,417	12	3	2.88	100
M (*SD*)		20 (3.7)	1,937.29 (365.71)	9.21 (2.8)		2.73 (0.27)	53.57 (29.03)
NBLI/FFW-L							
	1	22	2,161	10	1	2.33	13
	2	24	2,370	12	2	2.96	22
	3	24	2,392	12	2	2.96	54
	4	23	2,266	11	2	2.96	33
	5	22	2,189	12	2	2.96	98
	6	21	2,107	12	4	2.86	32
	7	19	1,893	11	4	2.86	73
M (*SD*)		22.14 (1.77)	2,196.86 (170.58)	11.43 (0.79)		2.84 (0.25)	46.43 (30.35)
Wait/NBLI							
	1	0	0	10	1	2.33	0
	2	0	0	10	1	2.33	0
	3	0	0	11	2	2.96	0
	4	0	0	9	2	2.96	0
	5	0	0	11	4	2.86	0
	6	0	0	12	4	2.86	0
	7	0	0	12	3	2.88	0
	8	0	0	12	3	2.88	0
	9	0	0	12	4	2.86	0
M (*SD*)		0	0	11 (1.12)		2.77 (0.25)	0

Attendance values for NBLI also are displayed in Table 2. Participants in the FFW-L/NBLI group attended an average of just over nine of the 12 scheduled treatment sessions. Participants in both the wait/NBLI and NBLI/FFW-L groups attended an average of approximately 11 sessions. The observed difference in attendance between the FFW-L/NBLI group and the NBLI/FFW-L group is statistically reliable, $t(12) = -2.02$, $p = .07$, but that between FFW-L/NBLI and the wait/NBLI group narrowly missed statistical significance, $t(14) = -1.76$, $p = .101$.

NBLI was administered by three doctoral students in speech-language pathology whose distribution across participants is displayed in Table 2. In 2006, Clinician 1 was ASHA certified, but Clinician 2 had not yet fulfilled her clinical fellowship requirements. In 2007, Clinician 2 was unavailable, so she was replaced by Clinician 3—an ASHA-certified student who had just begun doctoral training in speech-language pathology. Clinician 4 represents Clinician 1 in her 2nd year as an NBLI clinician.

To evaluate the fidelity of NBLI, each session was audio-recorded. After the interventions had concluded each summer, two graduate students in speech-language pathology who were not involved in the development of NBLI materials or in their administration were assigned to examine at least 50% of the audio-taped NBLI sessions. For each intervention session, the judges scored the clinician's adherence to the NBLI protocol using the sheet provided in the Appendix. If the clinician met all specified criteria, the maximum 23 points were recorded; otherwise, points were deducted for each component for which the clinician did not comply. In 2006, the mean ratings for the judges were 93.0% and 97.4% (*SD*s = 5.2% and 3.4%, respectively), with ranges from 82.61% to 100%. In 2007, mean judges ratings were 97.0% and 98.2% (*SD*s = 3.5% and 2.9%, respectively), with scores ranging from 91.3% to 100%. These scores indicate that the fidelity of NBLI administration was generally high. In 2006, however, they were notably lower for Clinician 1—an ASHA-certified doctoral student in speech-language pathology who served as a clinician in both years. Examination of the fidelity data revealed that only one judging category had sufficient variability within and across clinicians to reflect possibly real differences between clinicians. This category represents the judges' evaluations of the children's focus on intervention targets and the likelihood that problems with attention may have negatively influenced the clinician's ability to accomplish daily NBLI goals. A series of *t* tests indicated that Clinician 1's 2006 sessions were rated significantly lower in child attention and treatment focus than were those of Clinician 2, $t(44) = -4.63$, $p = .001$; Clinician 3, $t(32) = -2.86$, $p = .01$; and Clinician 4 (her own sessions in 2007), $t(28) = -2.28$, $p = .03$.

It is important to recognize that these lower ratings could be due to child factors or clinician factors or the interaction between the two. In any case, we attempted to discern the possibility that the small differences in treatment focus across clinicians could have biased the data in some way. The average treatment focus ratings assigned for each clinician at the time treatment was provided are 2.33, 2.96, 2.88, and 2.86 out of a total of three points for Clinicians 1, 2, 3, and 4, respectively. Table 2 provides the distribution of clinicians across children. It is evident from the table that Clinician 1's sessions were spread across the three treatment groups of 2006. A one-way ANOVA revealed that there were no significant differences in estimated treatment focus across the three treatment sequences, $F(2, 20) = 0.36$, $p = .71$. Thus, if there were systematic and meaningful differences in the clinician's handling of child compliance and behavior issues, they appear to have been distributed equally across the three experimental groups.

Statistical Analyses

For our two experimental questions, we first asked whether FFW-L leads to significant gains in language processing and use immediately after its administration (i.e., at 5 weeks; Time 2). More importantly, we asked whether it significantly boosts children's response to a 5-week conventional language intervention, NBLI, which focuses on narrative comprehension and production as well as complex grammar. The primary statistical tests for the first question involve the comparison of within-group gains from Time 1 to Time 2 and the comparison of group improvements by Time 2. The primary statistical tests for the second question require an examination of within-group changes from Time 1 to Time 3 and the comparison of group gains by Time 3, the first point at which all three groups had completed NBLI. We performed these tests with univariate repeated measures analyses via mixed model (McCulloch, 2005). To increase power in this exploratory study, we set the alpha for all comparisons at .10.

Results

The final column of Table 2 represents the percentage of game completion, averaged across all seven of the FFW-L games. As indicated in the table, most children in this study had difficulty progressing through the FFW-L games. Only one child in each group had a total completion rate of 90% or more and met SLC criteria for program completion. Seven children were able to complete an average of 50% or more of the program. The average percentage completed for the FFW-L/NBLI group ($M = 53.57$), however, was not significantly different

from the average for the NBLI/FFW-L group ($M = 46.43$), $t(12) = 0.45$, $p = .66$.

To address our two questions, we first tested the fixed effects in our mixed linear models for each of our four primary variables and for the two secondary variables. These tests were not significant for mean main verb scores on the DSS, $F(8, 36.91) = 0.88$, $p = .54$, or for mean conjunction use, $F(8, 35.17) = 1.50$, $p = .19$. The null hypothesis that the group means are the same across testing times could not be rejected for these variables. Therefore, we examined them no further in our focus on our two experimental questions.

In contrast, our mixed-model tests for fixed effects were statistically significant for two primary variables: the NRT, $F(8, 36.98) = 2.32$, $p = .04$, and the NLAI, $F(8, 33.58) = 3.17$, $p = .01$. The tests of fixed effects also were statistically reliable for the two secondary measures from the TNL: the Comp-raw, $F(8, 32.85) = 2.81$, $p = .02$, and the Oral-raw, $F(8, 34.93) = 2.48$, $p = .03$. These significant outcomes indicate that for each of these variables, we can reject the null hypothesis that all groups were the same at each testing point. Follow-up tests relevant to our experimental questions are appropriate for these variables.

The estimated marginal means for each group and time are presented for all measures in Table 3. D-Stat (Johnson, 1989) was used to calculate effect size, *d*, corrected for sample size, and 90% confidence intervals (CIs) around *d* for each contrast performed. These values are presented in Table 4.

Table 3. Estimated marginal means for all dependent variables across groups and times.

Measure	Group	Time 1	Time 2	Time 3
Mean main verb score	FFW-L/NBLI	2.20	2.12	2.37
	NBLI/FFW-L	2.45	2.69	2.38
	Wait/NBLI	2.32	2.43	2.35
Mean conjunctions per sentence	FFW-L/NBLI	0.12	0.09	0.14
	NBLI/FFW-L	0.13	0.17	0.17
	Wait/NBLI	0.18	0.20	0.24
NRT	FFW-L/NBLI	68.16_{a}	70.54	75.15_{a}
	NBLI/FFW-L	62.32_{b}	64.41	68.06_{b}
	Wait/NBLI	63.42	67.25	67.13
NLAI	FFW-L/NBLI	73.86_{a}	80.29_{a}	75.57_{e}
	NBLI/FFW-L	70.86_{bc}	76.00_{b}	80.29_{ce}
	Wait/NBLI	74.33_{d}	77.33_{f}	82.00_{df}
Comp-raw	FFW-L/NBLI	17.29_{ab}	20.86_{a}	19.86_{b}
	NBLI/FFW-L	18.57_{cd}	21.29_{c}	22.71_{d}
	Wait/NBLI	19.56_{e}	20.33	22.22_{e}
Oral-raw	FFW-L/NBLI	26.71	26.57	25.43_{de}
	NBLI/FFW-L	24.86_{a}	27.14	31.71_{ad}
	Wait/NBLI	25.11_{bc}	28.56_{b}	33.78_{bce}

Note. Within each measure, means with the same subscript represent statistically significant changes. Comp-raw = raw scores for the Narrative Comprehension composite of the TNL; Oral-raw = raw scores for the Oral Narration composite of the TNL.

Time 2: Isolated Treatment Effects

As shown in Tables 3 and 4, follow-up contrasts of effects for the NRT at Time 2 indicated that both the FFW-L/NBLI and NBLI/FFW-L groups made small, nonsignificant pre–post gains in their repetition of nonsense words (ts < 0.84, ps > .40). Furthermore, neither treatment group exhibited improvements that were significantly different from each other or from the wait/NBLI group, which served as a no-treatment control over the first 5 weeks of the experiment (ts < –0.44, ps > .66, ds < –0.21).

Planned comparisons indicated that NLAI scores for the FFW-L/NBLI group increased significantly from Time 1 to Time 2, $t(46.00) = 2.23$, $p = .03$, as did those for the NBLI/FFW-L group, $t(46.00) = 1.78$, $p = .08$. These effects were medium to large in size. In contrast, the wait/NBLI control group gains in NLAI at Time 2 were not statistically reliable, $t(40.00) = 1.18$, $p = .25$, $d = 0.35$. The strongest test of efficacy in our model, however, involves comparisons of the group gains after completion of the initial experimental treatment phase. Taken together, the gains in NLAI from Time 1 to Time 2 for the FFW-L/NBLI and the NBLI/FFW-L groups were significantly greater than those of the wait/NBLI group, $t(40.00) = 2.15$, $p = .04$, $d = 0.89$ (90% CI [0.15, 1.62]). Secondary analyses showed that the combined treatment groups outperformed the no-treatment wait group in narrative comprehension, $t(40.00) = 2.67$, $p = .01$, $d = 1.10$ (90% CI [0.35, 1.85]) but not in oral narration, $t(40.00) = –0.21$, $p = .83$, $d = –0.09$ (90% CI [–0.79, 0.62]). Taken separately, however, the NLAI improvements of the FFW-L/NBLI and the NBLI/FFW-L groups did not differ significantly from those of the wait/NBLI controls or from each other, ts(40.00) < 0.89, ps > .38, ds < 0.42.

Time 3: Effects of Sequenced Treatments

Planned comparisons at Time 3 revealed large, statistically significant gains from Time 1 to Time 3 on the NRT for the FFW-L/NBLI group, $t(38.00) = 2.45$, $p = .02$, and for the NBLI/FFW-L group, $t(38.00) = 1.86$, $p = .07$. In contrast, the mean NRT gain from Time 1 to Time 3 for the wait/NBLI group was not significant, $t(38.00) = 1.47$, $p = .24$. Despite the observed differences in within-group change over time, there were no differences between groups in the improvements exhibited from Time 1 to Time 3, ts(59.07) < 0.87, ps > .39, ds < 0.42.

Additional planned comparisons indicated that the mean NLAI scores for the FFW-L/NBLI group at Time 3 did not differ reliably from scores at Time 1, $t(40.00) = 0.59$,

Table 4. Pretest to posttest effect sizes and 90% confidence intervals (CIs) for all measures on which overall effects were observed.

		Time 1–Time 2		Time 2–Time 3		Time 1–Time 3	
Measure	**Group**	**Effect size**	**90% CI**	**Effect size**	**90% CI**	**Effect size**	**90% CI**
NRT	FFW-L/NBLI	0.27	[–0.61, 1.16]	0.53	[–0.36, 1.43]	0.81	[–0.11, 1.72]
	NBLI/FFW-L	0.22	[–0.66, 1.10]	0.39	[–0.50, 1.28]	0.61	[–0.29, 1.51]
	Wait/NBLI	0.46	[–0.33, 1.24]	–0.01	[–0.79, 0.76]	0.44	[–0.34, 1.23]
NLAI	FFW-L/NBLI	0.73	[–0.18, 1.64]	–0.54	[–1.43, 0.35]	0.20	[–0.69, 1.08]
	NBLI/FFW-L	0.59	[–0.31, 1.48]	0.49	[–0.40, 1.38]	1.07[a]	[0.13, 2.01]
	Wait/NBLI	0.35	[–0.43, 1.14]	0.55	[–0.24, 1.34]	0.91[a]	[0.09, 1.72]
Comp-raw	FFW-L/NBLI	0.86	[–0.05, 1.78]	–0.24	[–1.13, 0.64]	0.62	[–0.28, 1.52]
	NBLI/FFW-L	0.66	[–0.24, 1.57]	0.35	[–0.54, 1.23]	1.01[a]	[0.07, 1.94]
	Wait/NBLI	0.20	[–0.58, 0.97]	0.48	[–0.31, 1.26]	0.67	[–0.12, 1.47]
Oral-raw	FFW-L/NBLI	–0.02	[–0.90, 0.86]	–0.14	[–1.02, 0.74]	–0.15	[–1.03, 0.73]
	NBLI/FFW-L	0.27	[–0.61, 1.15]	0.54	[–0.35, 1.44]	0.81	[–0.10, 1.73]
	Wait/NBLI	0.42	[–0.36, 1.21]	0.64	[–0.15, 1.44]	1.07[a]	[0.24, 1.89]

[a]The lower boundary of the CI does not cross zero.

$p = .56$. In fact, the NLAI mean for this group was lower at Time 3 than at Time 2, although this decrement in performance was not statistically significant, $t(40.00) = -1.63$, $p = .11$. The gains between Time 1 and Time 3 for the NLAI were significant, however, for the NBLI/FFW-L group, $t(40.00) = 3.27$, $p = .002$, and for the wait/NBLI group, $t(40.00) = 3.01$, $p = .004$. As shown in Table 4, these effects were large, and the lower boundaries of the 90% CIs around these effects did not include zero, indicating a high probability of a significant positive change between Time 1 and Time 3 for these groups.

Again, however, the test of our central experimental hypothesis involves comparisons of the groups after each had completed 5 weeks of NBLI at Time 3. At this point, contrary to the experimental hypothesis, the change in NLAI of the FFW-L/NBLI group was reliably smaller than that of the NBLI/FFW-L group, $t(40.00) = -1.89$, $p = .07$, $d = -0.89$ (90% CI [–1.76, –0.03]). The mean NLAI change for the FFW-L/NBLI group also was lower than that of the wait/NBLI group, but this medium-to-large effect was not quite statistically significant, $t(40.00) = -1.55$, $p = .13$, $d = -0.74$ (90% CI [–1.59, 0.12]). However, the gains of the NBLI/FFW-L group did not differ significantly from the wait/NBLI controls, $t(40.00) = 0.46$, $p = .65$, $d = 0.22$ (90% CI [–0.61, 1.05]).

This pattern of results suggests that there could be unanticipated negative effects associated with completing NBLI after it follows an intensive course of FFW-L. To test this hypothesis, we first examined the changes in NLAI from the measurement point just before each group received NBLI to the test point immediately following NBLI. As noted previously, the FFW-L/NBLI group exhibited no improvements in NLAI from Time 2 to Time 3—the period during which they received NBLI. In contrast, both the NBLI/FFW-L group (from Time 1 to Time 2), $t(40.00) = 1.78$, $p = .08$, and the wait/NBLI group (from Time 2 to Time 3), $t(40.00) = 1.83$, $p = .07$, made significant, medium-sized gains in NLAI over their respective 5-week NBLI treatment periods. Next, we directly compared the changes in NLAI that occurred over the 5-week NBLI intervention period for the FFW-L/NBLI group (i.e., from Time 2 to Time 3) with the NLAI gains of the two groups that received NBLI as their first summer experimental intervention (i.e., NBLI/FFW-L from Time 1 to Time 2 and wait/NBLI from Time 2 to Time 3). This test was statistically significant, $t(40.00) = 2.77$, $p = .01$, $d = 1.21$ (90% CI [0.41, 2.01]), with a very large effect, indicating that children who received 5 weeks of NBLI before participating in any other summer language intervention improved their story skills (as measured by the NLAI) significantly more than did the children who received NBLI after the initial 5-week course of FFW-L.

The NLAI is the only norm-referenced measure we used as a dependent variable, and it is represented in standard score units that are referenced to the normative sample. Taking this into consideration, the gains of more than seven points by the wait/NBLI group and more than nine points by the NBLI/FFW-L group are large. With a standard error of measurement for the TNL of five standard score points (Gillam & Pearson, 2004), the 90% CI is + 8.25. Thus, the two-point average gain in NLAI over Time 1 to Time 3 made by the FFW-L/NBLI group falls well within the 90% CI around the Time 1 mean, whereas the nearly eight-point change of the wait/NBLI group falls just within the 90% upper limit, and the nine-point gain of the NBLI/FFW-L group exceeds the upper limit of the CI. An average gain of this magnitude for the NBLI/FFW-L group has less than a 10% probability of occurring by chance.

Secondary analyses based on the TNL. The estimated marginal means observed for Comp-raw and Oral-raw and the results of planned comparisons across Time and Group are displayed in Table 3. Associated pre–post effect sizes are presented in Table 4. Careful examination of the data for Comp-raw and Oral-raw clarifies the findings reported earlier on the NLAI in several ways. First, it is clear that the significant gain in NLAI from Time 1 to Time 2 for the FFW-L/NBLI group is based on a large improvement in narrative comprehension, $t(40.00) = 2.63$, $p = .01$. No gains were observed in oral narration over the same time period, $t(40.00) = 0.05$, $p = .96$. Second, the nearly statistically reliable decrement in NLAI for the same group between Times 2 and 3 was due to this group's failure to make gains in either narrative comprehension, $t(40.00) = -0.74$, $p = .47$, or oral narration, $t(40.00) = -0.41$, $p = .68$, over the period during which they received NBLI (i.e., Time 2 to Time 3). In sum, instead of an enhancement of learning ability evidenced by a large positive response to NBLI after first receiving a course of FFW-L, we observed a small, statistically negligible reduction in both Comp-raw and Oral-raw scores over this time period.

Third, the statistically significant gains in NLAI for the NBLI/FFW-L group from Time 1 to Time 3 resulted from medium-to-large Time 1–Time 3 gains in both Comp-raw, $t(40.00) = 3.06$, $p = .004$, and Oral-raw, $t(40.00) = 2.47$, $p = .02$. That is, the Time 1–Time 3 changes in NLAI observed for this group can be attributed to changes in both areas of narrative performance rather than to gains in one modality or the other. Similarly, the significant improvement in NLAI from Time 1 to Time 3 for the wait/NBLI group can be attributed to a medium-sized gain in Comp-raw, $t(40.00) = 2.23$, $p = .03$, along with a large gain in Oral-raw over the 10-week wait/NBLI experimental period, $t(48.38) = 3.54$, $p = .001$. Fourth, and finally, the only statistically significant between-groups effect observed for NLAI at Time 3 showed that the NBLI/FFW-L group had made significantly greater gains in NLAI than the FFW-L/NBLI group. The data in Tables 3 and 4 indicate that this effect was primarily attributable to the FFW-L/NBLI group's modest but, nonsignificant, decrements in performance between Time 2 and Time 3 for both Comp-raw and Oral-raw accompanied by a nearly significant increase in Oral-raw from Time 2 to Time 3 for the NBLI/FFW-L group, $t(40.76) = 1.65$, $p = .11$.

Supplementary analyses. It should be recalled that the FFW-L/NBLI group made significant strides in NLAI from Time 1 to Time 2, but the group's development then stopped and even appeared to regress slightly from Time 2 to Time 3—the time during which the group received NBLI. The groups differed minimally and nonsignificantly in their attendance of FFW-L sessions, but the FFW-L/NBLI group attended significantly fewer NBLI sessions than did the NBLI/FFW-L and wait/NBLI groups. Although this average difference is relatively small, it could explain the pattern of results for NLAI. Unfortunately, the three lowest NBLI attenders were in the FFW-L/NBLI group. These also were the only participants who did not attend at least nine NBLI sessions. This makes it impossible to compare directly the outcomes of the poor attenders in this group with those of equally poor attenders in the other groups.

Consequently, we examined the individual participant data to test the hypothesis that the unanticipated pattern for this group was caused more by the low NBLI attenders than the rest of the group. These data are presented in Table 5. As shown in the table, there is no support for the hypothesis. Both the low and high NBLI attendance subgroups made noteworthy contributions to the pattern of overall gains at Time 2, and both had one member whose NLAI decreased slightly from Time 1 to Time 2. Contrary to the tested hypothesis, two of the three low attenders remained stable or showed a gain from Time 2 to Time 3, whereas all four high attenders decreased in NLAI from Time 2 to Time 3. Thus, the overall group pattern of poor performance from Time 2 to Time 3 is driven primarily by the children who were high NBLI attenders, not low attenders.

If FFW-L has a positive impact on learning, one would expect that among those children who received FFW-L, there would be at least modest positive correlations between the number of minutes played or the percentage of programs completed and gains over Time 1–Time 3 on our outcome measures. Our correlational analyses rejected this expectation for NLAI, DSS main verb, conjunctions other than "and" and "then," and PPC on the NRT ($rs < -.28$, $ps > .34$, $N = 14$).

Discussion

Although other studies have examined the effects of FFW-L as an isolated treatment, none have explored its

Table 5. Observed NLAIs for participants in the FFW-L/NBLI group according to NBLI attendance patterns.

		NLAI		
Attender type	**Participant**	**Time 1**	**Time 2**	**Time 3**
Low NBLI attenders				
	1	85	100	94
	2	70	67	76
	3	64	67	67
M (*SD*)		73 (10.82)	78 (19.05)	79 (13.75)
High NBLI attenders				
	4	79	100	91
	5	61	64	55
	6	97	94	88
	7	61	70	58
M (*SD*)		74.5 (17.23)	82 (17.66)	73 (19.13)

potential as an adjuvant treatment designed to enhance children's ability to learn from another conventional language intervention. Despite its experimental structure, the present investigation was designed as a small, preliminary evaluation of FFW-L's capacity to do just that. Consequently, we had only secondary interests in and expected only limited effects of FFW-L as an isolated treatment on our measures of narration, grammar, and phonological memory. Instead, for our primary question, we asked, does participation in a 5-week course of FFW-L enhance the response of children with language impairment to a subsequent conventional narrative intervention? Our strongest and most important prediction was that the children in the FFW-L/NBLI group would outperform our other two groups on our outcome measures at Time 3 (especially those involving narration) after completion of both interventions in sequence.

Findings at Time 2

Time 2 was the only point in the study at which groups that received treatment could be compared with a no-treatment control group. As expected, treatment-related gains at this point were limited, and they were not reliably greater for either intervention group over the first 5-week period.

Grammatical expression and nonword repetition. On measures of grammatical expression collected during a conversational narrative with an unfamiliar adult and on a nonword repetition task, there were no significant effects either within groups over time or between groups. The lack of pre–post effects of FFW-L/NBLI on nonword repetition was somewhat surprising; this task was selected because it is believed to assess auditory attention and memory processes that are targeted directly by FFW-L.

Narrative expression and comprehension. Children in the FFW-L/NBLI group did exhibit medium-to-large, statistically significant pre–post gains on the NLAI, the major indicator of narrative performance from the TNL, at Time 2. Our secondary analyses demonstrated that gains in the NLAI for this group reflected significant improvements only in narrative comprehension. These pre–post gains in NLAI made by the children in the FFW-L/NBLI group by Time 2, however, were not significantly larger than those measured for the NBLI/FFW-L group, whose statistically significant NLAI gains at Time 2 also were specific to comprehension. It must be noted that neither treatment group on its own showed statistically significant improvement between Times 1 and 2 that was significantly greater than that made by the wait/NBLI group. However, there is evidence that these noneffects were the product of the low power in this preliminary study. When we compared the combined Time 1–Time 2 gains of the children who received either FFW or NBLI with the improvements of the children in the no-treatment wait group, the effect was large and statistically reliable. Thus, intervention had an immediate effect on narrative ability, especially narrative comprehension, but the effects of 5 weeks of FFW-L were not appreciably different from those of the far less intensive NBLI.

Comparison of results to those of other studies. Although the outcomes of a small study, such as ours, must be interpreted with caution, it is important to note that the patterns of results that we observed at Time 2 are generally consistent with the existing body of evidence on the effects of FFW-L reported by investigators not affiliated with the SLC. For example, the children with spoken and written language impairments studied by Pokorni et al. (2004) made no significant pre–post gains on the spoken language measures used in that study: three subtests from the Clinical Evaluation of Language Fundamentals, Third Edition (CELF–3; Semel, Wiig, & Secord, 1995). Thus, this study produced no signs that FFW-L facilitates spoken language skills. Other investigators—such as Cohen et al. (2005); Gillam et al. (2008); Hook, Macaruso, and Jones (2001); and Rouse and Krueger (2004)—noted gains immediately following FFW-L; however, as for our own findings, these gains were not significantly greater than those of other treatment or even control groups receiving no planned language interventions. Gillam et al. argued that even though the pre–post effects that they observed were not specific to FFW-L, they were sufficiently large to regard them as clinically significant. Regardless of one's position on this view, these and other findings presented by Gillam et al. indicate that the most innovative features of FFW-L—its acoustic modifications of the verbal and nonverbal stimuli—were not responsible for any gains in spoken language observed.

In contrast to these studies that generally have found no statistically significant treatment effects that are specific to FFW-L, Stevens et al. (2008) observed gains on the CELF–3 receptive language composite that were significantly greater than those made by a no-treatment control group. Differences between groups on the CELF–3 expressive language composite were not significant. These outcomes are remarkably similar to our own. A weakness of Stevens et al.'s study, however, is that the control group comprised children with typical language development. These children should not be expected to exhibit positive effects as large as those of children with language impairment, whose baseline scores are well below average. Stevens et al. presented electrophysiological evidence of improved selective attention after FFW-L, and this strengthens their claim of treatment efficacy. Because they evaluated no interventions other than FFW-L, however, it is impossible to determine whether this physiologic effect is specific to FFW-L or whether it could result from conventional treatments as well.

To summarize, neither FFW-L nor NBLI yielded significant improvements in children's grammatical production or ability to repeat nonwords. When they were grouped together and compared with children who received no intervention, participants who received 5 weeks of either FFW-L or NBLI exhibited significant improvements in narrative ability, principally in narrative comprehension. The gains of the children who participated in more than 30 hr of FFW-L, however, were not significantly larger than the changes of the children who received just less than 12 hr of NBLI. This pattern is not substantially different from that observed in other (quasi-)experimental studies published in peer-reviewed journals and conducted by researchers not associated with the SLC.

Findings at Time 3

Grammatical expression. For the grammar measures, no significant effects were observed, and examination of marginal means at each measurement point suggested that none of the intervention conditions yielded significant pre–post increases from Time 1 to Time 3. This result is disappointing because NBLI was created to have effects on children's grammatical expression. On the other hand, the finding is not entirely unexpected. In their evaluation of the feasibility of a version of NBLI that included 18 individual 1-hr sessions, Swanson et al. (2005) observed clear signs of NBLI effects based on a measure of oral narrative composition. However, they found no hint of effects on measures of spontaneous grammatical expression. After Swanson et al.'s study, we attempted to improve the NBLI stories and some of the components dealing with grammar, such as the sentence imitation stimuli. We also selected grammatical measures for the present study that were more specific to the children's individual NBLI goals than were those used by Swanson et al. Still, we offered only 12 group NBLI sessions in this study. Our expectation was that a boost provided by FFW-L to the FFW-L/NBLI group would yield significant effects on main verb and conjunction use in the conversational narrative context in which we collected our data, even with fewer sessions and less individual attention to each child than was provided in the original version of the intervention. However, the present study yielded no evidence of such a boost to grammatical ability or use.

It could also be that our measures of grammar, which were collected in a conversational narrative sample with an unfamiliar adult, simply were not sufficiently sensitive to change following a brief course of FFW-L and NBLI. There is independent evidence that is consistent with this hypothesis. In the only published FFW-L efficacy study to date to report measures of narration, Cohen et al. (2005) found that both the FFW-L and alternative computer instruction treatment groups and the no-treatment controls made significant improvements over time on a typical standardized test of grammar from the TOLD–P:3 (Newcomer & Hammill, 1997). In contrast, they observed no pre–post gains on more functional indices of grammatical development, sentence length and subordinate clause use from the Bus Story Test (Renfrew, 1997)—an assessment of children's narrative ability. If we had included additional, somewhat less functional measures of grammatical expression, we may have seen more positive signs of developmental change (for a relevant task alternative, see Gummersall & Strong, 1999).

Nonword repetition. In contrast to our measures of grammar, both groups that received FFW-L displayed statistically significant and medium-to-large changes in nonword repetition performance from Time 1 to Time 3. No such gains were observed for the wait/NBLI group. Thus, this effect appears to be specific to groups that received FFW-L along with NBLI. On direct comparison at Time 3, however, the groups that received FFW-L did not differ reliably from each other or from the wait/NBLI controls. The differences in effect sizes between the wait/NBLI group and the other two groups shown in Table 4 suggest that this nonsignificant finding could reflect the low power in our study. Thus, it may be that over the 11 weeks between our Time 1 and Time 3 testing sessions, children who received FFW-L and our conventional narrative treatment improved significantly in their ability to repeat nonwords. It is difficult to know what to make of this finding because nonword repetition is, in itself, not a particularly functional skill. On the other hand, the NRT is generally taken to be a processing-dependent rather than a knowledge-dependent measure (Campbell, Dollaghan, Needleman, & Janosky, 1997). In other words, the combination of FFW-L and NBLI, especially with FFW-L first, might have facilitated children's development of a set of underlying phonological processes that are important to language learning. Because neither the FFW-L/NBLI group nor the NBLI/FFW-L group showed statistically reliable gains in NRT after only 5 weeks of their intervention, and the wait/NBLI group received only 5 weeks of intervention, we cannot rule out the alternative possibility that effects on NRT performance were due simply to time spent in intervention. Because we had no further assessments of the children's progress, we cannot evaluate this possibility further.

Narrative expression and comprehension. Following their significant improvement in narration after a 5-week course of FFW-L, children in the FFW-L/NBLI group made no subsequent gains on our key measure of narrative performance, the NLAI, from Time 2 to Time 3 in response to NBLI. In contrast, both of the comparison groups made medium-sized gains in NLAI between Time 2 and Time 3. The average Time 1–Time 3 gain of the

NBLI/FFW-L group even exceeded the 90% confidence limits based on the published standard error of measurement estimates for the TNL.

How can our FFW-L/NBLI group's unpredicted, especially poor response to NBLI be explained? We have no fully satisfactory answer to this question, but one reasonable speculation is that the children in this group may simply have "burned out" over the course of their FFW-L experience. This may have left them less well prepared to learn from NBLI than were the children in the other two groups, both of whom received NBLI as their first project intervention. We cannot determine definitively that the children in our FFW-L/NBLI group burned out or developed attitudes that may have limited their learning in response to NBLI or their performance on our Time 3 test battery. Other users of FFW-L, however, have reported observations consistent with this speculation. For example, we encountered a relevant citation on the SLC Web site in which a teacher attempted to relieve stress among FFW-L users by adopting one of the alternative protocols (e.g., 50 min per day for 12 weeks) now made available by the SLC:

> Phipps cautions that it is imperative to use the right protocol to get optimal results. "The first year, we were running a 100-minute protocol, and the kids were just burning out after a while. But then we switched to the 40- and 50-minute protocol, and the kids are enjoying it. They think they are playing games," he says, "and they just don't realize how much they're learning." (SLC, 2008, para. 10)

There are several factors that could have increased the likelihood that burnout was a factor in our application of FFW-L for the FFW-L/NBLI group. First, the challenges that FFW-L posed for our participants were significant, with neither group averaging more than 55% completion of program games. Our children's experience in this regard was not unique; in fact, most studies to date have commented on the general difficulties children have had playing many of the games and the relatively small numbers of children who reached SLC program completion standards, even when language outcomes have been positive (see, especially, Stevens et al., 2008).

Second, our service delivery model for FFW-L may have increased stress compared with participants in other studies of FFW-L. Our FFW-L participants arrived at 9:00 a.m., and our sessions concluded promptly at 11:00 a.m. Besides the 100 min for FFW-L games, there were only 20 min left for snacks and for the children to total their points, check to see whether they had met their goals, and choose the rewards that they had earned for the day. Despite clinicians' efforts to interact in a warm and friendly way with all children at all times, it is clear that this schedule left little time for play with other types of games, real summertime fun, or meaningful interactions with adults and peers that could have alleviated stress and provided further opportunities for socialization and language learning. By comparison, the children participating in the FFW-L group in the study by Gillam et al. (2008), along with their concentrated 100 min of FFW-L, "also received a 20-min snack break and attended a 1.5-hr 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)" (p. 103). This may have helped to limit stress levels and to provide more opportunities to learn from other sources compared with our story camp.

Note that it cannot be argued that the children in the NBLI/FFW-L group faced similar challenges and schedules and, thus, also should have shown a progress slowdown or regression during their second treatment period, during which they received FFW-L. These children attended only 2.5 NBLI sessions per week during their first 5 weeks of story camp, and although they were at camp for a full 2 hr, the structured NBLI activities were sandwiched between snacks, listening to books, drawing, and other activities. Although NBLI intervention procedures were avoided during these activities, they offered many opportunities to relax and have fun and to interact with clinicians and peers on each NBLI intervention day.

Study Limitations

It is important to note several additional study limitations and to examine the extent to which they serve as threats to the internal and external validity of our investigation. First, and perhaps most obviously, our participant sample was small. The most significant internal validity problems associated with a small *N* generally are related to low statistical power and the threat of Type II errors. We addressed this problem from the outset by raising alpha to .10. In addition, when it was meaningful to do so, we combined treatment groups for specific group contrasts, thus, increasing sample size and power. Nevertheless, with such a small *N,* our statistical design was sensitive only to relatively large effects, and caution is required in interpretation of our findings.

A second shortcoming of our study that is related to the small size of our sample is that seven of 30 (23%) children who had been identified and enrolled for the study either withdrew from the study shortly after their Time 1 testing (five children) or were excluded from the study because they failed to participate in at least half of all of their scheduled sessions (two children). Five of these seven children were lost to follow-up. With such a small *N,* we elected to exclude these children from all analyses rather than to use an intention-to-treat methodology. We presented the preexperimental and Time 1 data for these children in Table 1, however, to enable comparisons with the participating groups. There were

no clear-cut differences between these children and those who were included in all analyses, and there were not obvious differences in the excluded children who had been assigned to the FFW-L/NBLI group versus the NBLI/FFW-L group. Therefore, it seems unlikely that the relatively large number of withdrawals could have led to the pattern of noneffects and negative effects that we found to be associated with the FFW-L/NBLI treatment sequence.

A third limitation of the study is closely related to the first two. Of the children who were included in the study, several children had poor attendance records. It is possible that some children's responses to treatment were mitigated by the relatively low intensity of their treatment regimens (Warren, Fey, & Yoder, 2007). Evidence both internal and external to our study, however, suggests that low-treatment intensity is not an adequate explanation for our results. For example, we found low, nonsignificant correlations between performance on our outcome measures and days of FFW-L intervention attended, minutes of games played, and average percentage of game completion. This finding is not anomalous. Cohen et al. (2005), Hook et al. (2001), Rouse and Krueger (2004), and Stevens et al. (2008) each examined the relationship between exposure to FFW-L games and/or the level of completion of FFW-L goals and outcome measures and, with few small exceptions (see Cohen et al., 2005), found nonsignificant correlations. The participants in Stevens et al.'s study completed lower percentages of FFW-L games than ours did, yet they still made significant improvements in receptive language. These investigators interpreted their findings to suggest "that something other than progression through the FFW exercises may be responsible for the receptive language gains observed, for example, the daily interactions with the FFW coaches or improved attention" (Stevens et al., 2008, p. 63). On the basis of all of the evidence available, this is a reasonable possibility worthy of future consideration.

Summary and Conclusions

We undertook this investigation as a preliminary test of the hypothesis that FFW-L may be effective as an adjuvant treatment, enhancing the effects of conventional language interventions, even if it is not found to be efficacious as an isolated intervention. As we have stressed throughout this article, the findings of our small study must be interpreted cautiously. Bearing this in mind, our results provide no support for our primary hypothesis. Therefore, on the basis of our outcomes and others we have reviewed, our study does not provide us with the motivation to continue our plans to conduct a major trial of FFW-L's effects on the language-learning capabilities of children with spoken language impairments. Instead, studies of FFW-L in the immediate future should be small and driven by highly specific hypotheses. For example, on the basis of our results and speculations, it would be sensible to compare directly the effects of the findings of FFW-L when presented in one of the shorter protocols now made available by the SLC compared with the original 100-min regimen. Additional studies probing the mediating effects of FFW-L on children's working memory and related language skills would also be in order. As another example, the findings of Stevens et al. (2008) might lead researchers to evaluate the effects of FFW-L on children with spoken language impairments who have identified problems with selective attention.

From a clinical perspective, when considered alongside the available scientific evidence on FFW-L, our study provides little justification for implementing FFW-L as a supplement to conventional interventions for children with spoken language impairments. If it is chosen as a primary or supplementary language intervention, clinicians might consider alternative, less-intensive, service-delivery models—such as the 50- and 75-min versions—to limit stress and reduce the possibility of participant burnout. Whether or not this step is taken, clinicians selecting FFW-L should make it routine clinical practice to inform the parents of children scheduled to receive it that there is a substantial amount of evidence that, at best, the program's outcomes should be expected to be no greater than equally intensive, conventional language interventions.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders (NIDCD) Grant R21DC007214. Additional support was provided by NIDCD Grant P30DC005803 and National Institute on Child Health and Human Development (NICHD) Grants P30HD02528, T32HD007489, and P30HD003352. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDCD, the NICHD, or the National Institutes of Health. We thank Sophie Ambrose, Stephanie Becker, Nadine Devlin, Elizabeth Elliott, Katie Irwin, Britney Jurgensen, Leslie Lefebvre, Steven Long, Josie Row, Mindy Sittner, Shari Sokol, and Peggy Waggoner for their assistance in providing intervention, diagnostic, transcription, and coding services. We are especially grateful to Cindy Lane, Claudia Shannon, Dan Wright, and the speech-language pathologists of U.S. District 500 for their roles in helping us to identify candidate participants for the study.

References

Agocs, M. M., Burns, M., DeLey, L. E., Miller, S. L., & Calhoun, B. M. (2006). Fast ForWord Language. In R. J. McCauley & M. E. Fey (Eds.), *Treatment of language disorders in children* (pp. 471–508). Baltimore, MD: Brookes.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173.

Cable, A. L. (2007). *An oral narrative intervention for second graders with poor oral narrative ability*. Unpublished doctoral dissertation, University of Texas at Austin.

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language, and Hearing Research, 40,* 519–525.

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45,* 1142–1157.

Cirrin, F. M., & Gillam, R. B. (2008). Language intervention practices for school-age children with spoken language disorders: A systematic review. *Language, Speech, and Hearing Services in Schools, 39,* S110–S137.

Cohen, W., Hodson, A., O'Hare, A., Boyle, J., Durrani, T., McCartney, E., & Watson, J. (2005). 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, 48,* 715–729.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1136–1146.

Ellis Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 43,* 865–878.

Fey, M. E., Catts, H. W., & Larrivee, L. S. (1995). Preparing preschoolers for the academic and social challenges of school. In M. E. Fey, J. Windsor, & S. F. Warren (Eds.), *Language intervention: Preschool through the elementary years* (pp. 3–38). Baltimore, MD: Brookes.

Fey, M. E., Cleave, P. L., Long, S. H., & Hughes, D. L. (1993). Two approaches to the facilitation of grammar in children with language impairment: An experimental evaluation. *Journal of Speech and Hearing Research, 36,* 141–157.

Finestack, L. H., Fey, M. E., Sokol, S. B., Ambrose, S., & Swanson, L. A. (2006). Fostering narrative and grammatical skills with "syntax stories." In A. V. Kleeck (Ed.), *Sharing books and stories to promote language and literacy* (pp. 319–346). San Diego, CA: Plural.

Gillam, R. B., Loeb, D. F., Hoffman, L. M., Bohman, T., Champlin, C. A., Thibodeau, L., & Friel-Patti, S. (2008). The efficacy of Fast ForWord Language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 51,* 97–119.

Gillam, R., McFadden, T. U., & van Kleeck, A. (1995). Improving narrative abilities: Whole language and language skills approaches. In M. E. Fey, J. Windsor, & S. F. Warren (Eds.), *Language intervention: Preschool through the elementary years* (pp. 145–182). Baltimore, MD: Brookes.

Gillam, R. B., & Pearson, N. A. (2004). *Test of Narrative Language*. Austin, TX: Pro-Ed.

Goldman, R., & Fristoe, M. (2000). *Goldman Fristoe Test of Articulation–Second Edition*. Circle Pines, MN: AGS.

Gummersall, D. M., & Strong, C. J. (1999). Assessment of complex sentence production in a narrative context. *Language, Speech, and Hearing Services in Schools, 30,* 152–164.

Hook, P. E., Macaruso, P., & Jones, S. (2001). Efficacy of Fast ForWord training on facilitating acquisition of reading skills by children with reading difficulties—A longitudinal study. *Annals of Dyslexia, 51,* 75–96.

Johnson, B. T. (1989). *Software for the meta-analytic review of research literatures*. Mahwah, NJ: Erlbaum.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test, Second Edition*. Circle Pines, MN: AGS.

Law, J., Garrett, Z., & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 47,* 924–943.

Leadholm, B. J., & Miller, J. F. (1992). *Language sample analysis: The Wisconsin guide*. Madison, WI: Wisconsin Department of Public Instruction.

Lee, L. L. (1974). *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.

Leonard, L. B., Camarata, S. M., Brown, B., & Camarata, M. N. (2004). Tense and agreement in the speech of children with specific language impairment: Patterns of generalization through intervention. *Journal of Speech, Language, and Hearing Research, 47,* 1363–1379.

Loeb, D. F., Stoke, C., & Fey, M. E. (2001). Language changes associated with Fast ForWord-Language: Evidence from case studies. *American Journal of Speech-Language Pathology, 10,* 216–230.

McCulloch, C. E. (2005). Repeated measures ANOVA, RIP? *Chance, 18,* 29–33.

McGregor, K. K. (2000). The development and enhancement of narrative skills in a preschool classroom: Towards a solution to clinician–client mismatch. *American Journal of Speech-Language Pathology, 9,* 55–71.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996, January 5). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science, 271,* 77–81.

Miller, J. F., & Chapman, R. (2000). SALT: Systematic Analysis of Language Transcripts [Computer software]. Madison: University of Wisconsin—Madison, Waisman Center, Language Analysis Laboratory.

Nelson, K. E., Camarata, S. M., Welsh, J., Butkovsky, L., & Camarata, M. (1996). 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 and Hearing Research, 39,* 850–859.

Nelson, N. W., & Hyter, Y. (1990, November). *Black English Sentence Scoring: Development and use as a tool for non-biased assessment*. Short course presented at the Annual Convention of the American Speech-Language-Hearing Association, Seattle, WA.

Newcomer, P. L., & Hammill, D. D. (1997). *Test of Language Development–Primary, Third Edition*. Austin, TX: Pro-Ed.

Oetting, J. B., & Cleveland, L. H. (2006). The clinical utility of nonword repetition for children living in the rural south of the U.S. *Clinical Linguistics & Phonetics, 20,* 553–561.

Pokorni, J. L., Worthington, C. K., & Jamison, P. J. (2004). Phonological awareness intervention: Comparison of Fast ForWord, Earobics, and LiPS. *Journal of Educational Research, 97,* 147–157.

Popescu, M., Fey, M. E., Lewine, J. D., & Popescu, E. A. (2009). N400 responses of children with primary language disorder: Intervention effects. *Neuroreport, 20,* 1104–1108.

Renfrew, C. (1997). *Bus Story Test* (4th ed.). Bicester, England: Winslow Press.

Rouse, C. E., & Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a "scientifically based" reading program. *Economics of Education Review, 23,* 323–338.

Scientific Learning Corporation. (2008). *Cumberland County, NC*. Retrieved from www.scilearn.com/results/success-stories/case-studies/cumberl-county-school-district-sees-amazing-results-for-all-types-of-students-using-fast-forwordsup-sup-software.php.

Scientific Learning Corporation. (2009). *Scientific learning products*. Retrieved from www.scilearn.com/products/index.php?source=Google&adgroup=ffw3 9-22-8.

Semel, E., Wiig, E. H., & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals, Third Edition*. San Antonio, TX: The Psychological Corporation.

Stark, R. E., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders, 46,* 114–122.

Stevens, C., Fanning, J., Coch, D., Sanders, L., & Neville, H. (2008). Neural mechanisms of selective auditory attention are enhanced by computerized training: Electrophysiological evidence from language-impaired and typically developing children. *Brain Research, 1205,* 55–69.

Swanson, L. A., Fey, M. E., Mills, C. E., & Hood, L. S. (2005). Use of narrative-based language intervention with children who have specific language impairment. *American Journal of Speech-Language Pathology, 14,* 131–141.

Tallal, P. (2000). Experimental studies of language learning impairments: From research to remediation. In D. V. M. Bishop & L. B. Leonard (Eds.), *Speech and language impairments in children* (pp. 131–155). Baltimore, MD: Brookes.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., & Merzenich, M. M. (1996, January 5). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science, 271,* 81–84.

Thibodeau, L. M., Friel-Patti, S., & Britt, L. (2001). Psychoacoustic performance in children completing Fast ForWord training. *American Journal of Speech-Language Pathology, 10,* 248–257.

Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The stability of primary language disorder: Four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research, 46,* 1283–1296.

Ukrainetz, T. A. (1998). Stickwriting stories: A quick and easy narrative representation strategy. *Language Speech and Hearing Services in Schools, 29,* 197–206.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13,* 70–77.

Westerveld, M. F., & Gillon, G. T. (2008). Oral narrative intervention for children with mixed reading disability. *Child Language Teaching & Therapy, 24,* 31–54.

Received October 28, 2008

Revision received May 14, 2009

Accepted July 6, 2009

DOI: 10.1044/1092-4388(2009/08-0225)

Contact author: Marc E. Fey, Department of Hearing and Speech, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160-7605. E-mail: mfey@kumc.edu.

Lizbeth H. Finestack is now at the University of Minnesota.

Appendix. Fidelity of treatment for narrative-based language intervention.

Clinician:	**Session number:**	**Session date:**	**Treatment period:**
Participant #s:	**Participant goals:**	**Session time:**	

	Possible	Observed
1. Did the session begin with repetition of a story from the previous week (except Session 1)?	4	0
a. Was each part of the story reviewed with story keys by asking children to recall details?	1	
b. Did each participant have an opportunity to recall details from at least one part of the story?	1	
c. Did the clinician adequately scaffold child attempts at the recalling story details?	1	
d. Did the clinician then reread the complete version of the story from the previous week?	1	
2. Did the session proceed with retell of a new story?	5	0
a. Did the clinician ask for child comments and questions on the story topic?	1	
b. Did the clinician use the story keys to stress each part of the story?	1	
c. Did each child have an opportunity to repeat two parts of the story (or 1 in 1 or 2 action stories)?	1	
d. Was each part of the story represented in the children's collaborative retell?	1	
e. Did the clinician adequately scaffold child attempts at the retell?	1	
3. Did each child imitate 4–12 sentences containing a target grammatical structure?	3	0
a. Did the clinician properly reinforce the child for imitating the target?	1	
b. Did the clinician simplify or break down the target when the child experienced difficulty?	1	
c. If given the opportunity, did the clinician recast inaccurate repetitions of the target?	1	
4. Did the session proceed with generation of a new story?	7	0
a. Did the clinician help the children to negotiate a problem and beginning for the story?	1	
b. Did the clinician use the keys to help the children to represent each story part?	1	
c. Did each child have an opportunity to generate at least two parts of the story?	1	
d. Did the clinician tell a coherent rendition of the total story?	1	
e. Was each part of the story represented in the children's retell?	1	
f. Did the clinician adequately scaffold child attempts at the retell?	1	
g. Did the clinician provide an opportunity to tell the story to an uninformed listener?	1	
5. Did the clinician produce at least five target-specific recasts for each child?	1	
6. Did the clinician maintain the children's focus on treatment activities in a way that appeared to optimize achievement of teaching goals?	3	
	Total	Total
Key: 3 = High level of child attention	23	0
2 = Some concerns about child attention		
1 = Clear concerns about attention across the session		
0 = Attention is so poor goals cannot be reached		