RUNNING HEAD: Fast ForWord Randomized Field Trial

Can Brain Research and Computers Improve Literacy?

A Randomized Field Trial of the Fast ForWord Language Computer-Based Training Program

Can Brain Research and Computers Improve Literacy?

A Randomized Field Trial of the Fast ForWord Language Computer-Based Training Program

Abstract

This article describes an independent assessment of the Fast ForWord Language computer-based training program developed by Scientific Learning Corporation. Previous laboratory research involving children with language-based learning impairments showed strong effects on their abilities to recognize brief and fast sequences of nonspeech and speech stimuli, but generalization of these effects beyond clinical settings and student populations and to broader literacy measures remains unclear. Implementing a randomized field trial in eight urban schools, we generated language and reading comprehension impact estimates for second- and seventh-grade students more generally at-risk for poor literacy outcomes. There were some problems of implementation in the field setting and intention-to-treat effects, $d = .21$, were restricted to one outcome, reading comprehension, for seventh grade students assigned to treatment. Supplemental analyses, which rescaled the intention-to-treat impacts by the average program completion rate, also revealed reading comprehension effects of the treatment on the treated for seventh graders.

Drawing on 30 years of research on brain plasticity and the neurological basis of language and literacy, scientists developed the Fast ForWord computerized educational products in an effort to provide training in oral language comprehension and other skills critical for success in reading. The programs are based on the theory that many children experiencing reading and language difficulties have a general auditory processing deficit that stems from their difficulties detecting rapidly presented and changing stimuli. Phonics, or the elements of sound called phonemes that make up words, may be broken down even further to smaller phonetic elements or acoustic events. These elements are rapidly, and rather automatically, processed by most people as they hear syllables, words, and sentences articulated. However, as Fitch & Tallal (2003) argued, disturbances in auditory processing may contribute not only to problems in speech perception but also to weaknesses in developing phonemic awareness skills and learning and applying phonic reading strategies.

For instance, though consonant-vowel pairs (e.g., /ba/ and /da/) mainly differ in the first 40 ms of the speech signal (Reed, 1989), children experiencing language difficulties require hundreds of milliseconds to discriminate such speech signals (Tallal, Miller, Jenkins, & Merzenich, 1997). As a result, these children process the rapid and successive incoming sensory information too slowly to discriminate among the brief and weak acoustic cues that are needed to learn the phonemes of language. This impairment may then impede a child's ability to understand and use language, read, perform at a satisfactory level in school, and participate successfully in many aspects of the home and school environment (Benasich, Thomas, Choudhury, & Leppanen, 2002; Tallal, 1980).

The important link between understanding the sounds of language, or phonological awareness, and children's acquisition of broader reading skills is supported by ample empirical evidence (Adams, 1990; Wagner & Torgesen, 1987). Indeed, there is strong consensus that phonological awareness is the best single predictor of reading ability, not just in the early grades (Ehri & Wilce, 1980; 1985; Perfetti, Beck, Bell, & Hughes, 1987) but throughout the school years (Calfee, Lindamood, & Lindamood, 1973; MacDonald & Cornwall, 1995; Shankweiler et al., 1995). Research demonstrates that students without this awareness will encounter difficulties when learning to read. In order to have phonological awareness, though, students must be able to accurately process the incoming sounds of language. The exercises in the Fast ForWord products are designed to break down the components of language by altering the way the brain processes the rapidly successive features of speech sounds. In addition to work on the rate of auditory processing, students work systematically with phoneme identification and discrimination, sound matching and auditory word recognition. Based on the arguments of one of the program's developers, Paula Tallal (1980), the programs help facilitate learning of both language and reading because they sharpen phonological processing skills—including speech perception and phonemic awareness—which in turn benefit acquisition of phonic word attack strategies.

Recent data from the Nation's Report Card, the National Assessment of Educational Progress (NAEP), indicated that more than two thirds of the nation's fourth graders (69%) and eighth graders (67%) scored below the level of "proficient" for their grade level in 2002 (Grigg, Daane, Yin, & Campbell, 2003). Although Fast ForWord was originally designed in 1996 for children with language-based impairments, over the years the program has been marketed to and applied to this much larger and increasingly diverse group of students in the United States who are more generally having trouble reading. There have been claims of considerable success documented in the popular media, as suggested by Tallal, who was quoted in a recent *Newsweek* article as stating that "...90 percent of the kids who complete the program made 1.5 to two years of progress in reading skills" (Begley, 2000, p. 64). However, beyond initial laboratory experiments involving children with specific language-based impairments, there has been little rigorous research to assess the program's effects for the larger population of struggling readers.

This article describes the methods and results of an independent assessment of the Fast ForWord Language computer-based training program developed by Scientific Learning Corporation. Working in eight Baltimore City Public Schools, we contrasted the language and reading outcomes for second and seventh grade students who received the supplemental Fast ForWord program, and their regular classroom-based literacy instruction, to the outcomes for control students who received only the regular classroom instruction and no supplemental services. Data collected included a pretest and posttest of students' language and reading comprehension skills using the Comprehensive Test of Basic Skills, Fifth Edition (CTBS/5) Terra Nova, teacher-reported information regarding each child's language skills prior to the intervention, and descriptive data regarding the implementation of the program at the school sites. We implemented an experimental design that involved within-school random assignment of students to the program or a control group condition. This randomized field trial across eight Baltimore schools yielded a study with strong internal validity for assessing whether the Fast ForWord Language programs can be expected to help academically at-risk students from an urban locale learn literacy skills that are commonly measured in school accountability programs across the nation.

Background

Using computer technology, the Fast ForWord programs' mathematical algorithms change many of the features of oral language, including volume, pitch, and duration. The programs selectively stretch and emphasize important acoustic differences in the sounds of language so that individuals recognize the acoustic “signature” of each phoneme. Both the stretching and the enhancement are varied systematically in degree and intensity over the course of training. The programs are adaptive, in that they responds to each participant’s incoming skill level and ongoing progress by gradually moving speech sounds closer and closer together with less and less modification until the student is eventually learning with more natural speech acoustics.

The family of Fast ForWord products includes three programs: Fast ForWord Language (available in both elementary and middle/high school versions), Fast ForWord Language to Reading, and Fast ForWord Reading. Fast ForWord Language is designed to build oral language comprehension skills such as phonemic awareness, auditory processing speed, phonological awareness, working memory, syntax, grammar, sequencing and other critical skills necessary to learn to read or to become a better reader. The elementary and middle/high school variants of the Fast ForWord Language program are similar, but the middle/high school software contains graphics and content tailored to older students. Language to Reading focuses on making the connection between spoken and written language and attends to skills such as sound/letter recognition, decoding, vocabulary, syntax and grammar, listening comprehension, and word recognition. Finally, Fast ForWord Reading helps children build reading skills such as word recognition and fluency, decoding, vocabulary, spelling, and passage comprehension.

The Fast ForWord Language program, which is the focus of this study, consists of 100 minutes of training per day, 5 days a week, for 4 to 8 weeks under the supervision of Fast

ForWord-trained clinicians or educators.[1] The developers require this high degree of intensity and frequency to sustain a level of practice that develops skills to the extent that they become automatic, or learned. The exercises are provided in a game-like computerized environment, with animations to help maintain the child's interest. On-screen rewards for successful completion of training segments are supplemented with token economy rewards, which are awarded for achievement of point goals determined in conjunction with the participant. Points are awarded for a combination of correct answers and attentiveness to instruction. Data from the child's daily exercises are uploaded via an Internet connection to Scientific Learning, and weekly reports are generated to summarize the child's progress. Upon entering the program, all children start at the same basic level and advance only after attaining a predetermined level of proficiency. The rate at which a child progresses through the program is, therefore, determined by the child. On average, children complete Fast ForWord Language in 6 to 8 weeks.

The Fast ForWord Evidence Base

The initial research on the prototype that would, ultimately, become the Fast ForWord Language program, revealed that children with language-based learning impairments (i.e., major deficits in their recognition of some rapidly successive phonetic elements and non-speech stimuli) improved markedly in their abilities to recognize brief and fast sequences of nonspeech and speech stimuli with approximately 20 hours of computer-based training per week over a four-week period (Merzenich et al., 1996; Tallal et al., 1996).

Additional research used functional magnetic resonance imaging to analyze the brain activity of 20 dyslexic children and 12 children with normal reading abilities while they performed a simple rhyming task (Temple et al., 2003). After the initial scans, the 20 dyslexic

[1] The Fast ForWord Language program for middle and high-school students has a recommended schedule of 90

children used the Fast ForWord Language training program for eight weeks as part of their regular school day. At the end of the eight weeks of training, brain scans revealed increased activity in the brains of the dyslexic children that represented a partial amelioration of the disrupted responses found prior to Fast ForWord training. The authors, who also found improved language and reading test scores for the Fast ForWord group, concluded that the program may have changed brain areas related to the sound structure of language, which in turn led to improved language and reading outcomes.

These early studies yielded impressive results suggesting that Fast ForWord Language could hold considerable promise for improving students' language and reading outcomes. However, the ability to generalize the effects from these small and tightly controlled laboratory experiments to typical implementations of Fast ForWord Language in school-based settings may be limited. Cronbach et al. (1980) referred to the superrealization stage of program development to characterize such small-scale evaluations of promising new programs implemented under optimal conditions. These types of efficacy studies demonstrate the optimal effects that can be expected of an intervention, but generally do not provide realistic estimates of what can be expected from an intervention when it is implemented on a broader scale in schools, classrooms, and other field-based settings.

Further, the applicability of these findings may be modest, because these studies generalize from very small samples of children who have specific language-based impairments that occur at a rate of only about 7% to 8% in the U.S. population (Fitch & Tallal, 2003). Nevertheless, in the years since this initial research, the developers have disseminated and applied Fast ForWord Language to an expanded range of other student subgroups beyond the

---

minutes per day, 5 days per week, for a minimum of 20 days.

specialized groups of language-impaired children originally targeted by the intervention. These groups have included children with developmental problems or classifications, such as Pervasive Developmental Disorder (PDD), Attention Deficit Disorder (ADD), and Central Auditory Processing Disorder (CAPD) (Gillam, 1999). Further, various pretest-to-posttest outcome data presented at the developers' website suggest that a variety of other children may benefit, including students at risk for reading failure, students from low socioeconomic status backgrounds, at-risk African American students, and English language learners (see www.scientificlearning.com/scie/index.php3?main=graphs& cartid=).

These data and other information provided by the Fast ForWord website suggest that a broad school-based population of "…students who are struggling with reading and need to develop the cognitive skills necessary for successful reading and learning" can benefit from the Fast ForWord Language products (see http://www.scilearn.com/prod2/ffwd_l/main=home). Yet, only one known randomized field trial has assessed the impacts of Fast ForWord Language in school-based settings with a more general sample of students struggling with reading. Rouse and Krueger (2004) found that the programs improved some aspects of students' language skills, but they did not find that these gains translated into impacts on broader measures of language acquisition or actual reading skills, at least as measured by commonly used standardized tests. Though Rouse and Krueger acknowledged that their sample sizes were too small to detect small effects of Fast ForWord Language of less than 0.10 *SD*s with statistical precision, they also noted that such impacts are considerably smaller than those of approximately 0.80 *SD*s that have been promoted by the developers.

Therefore, there is some anecdotal and non-experimental information from the developer that suggests that a wider variety of students than previously assumed may achieve important improvements on a range of literacy skills as a result of training with Fast ForWord Language. Only Rouse and Krueger (2004), though, have provided rigorous experimental evidence concerning such effects. In this respect, the true impacts of Fast ForWord Language implemented in school-based settings and targeting students generally at risk for poor literacy performance do not appear to be well understood.

Objectives and Hypotheses

The current study examined the overall effects of the Fast ForWord Language program on at-risk students' reading and language outcomes. The children participating in the study were predominantly African American, of low socioeconomic status, and were performing below national averages on norm-referenced reading achievement tests. In addition, the study investigated two potential aptitude-by-treatment (ATI) interaction effects related to the following baseline student characteristics:

- The degree of pre-intervention speech and language difficulties experienced by the child; and
- The extent of pre-intervention reading problems experienced by the child.

That is, do students with more severe language and speech problems and/or those with more difficulty with reading realize stronger language and reading achievement gains from Fast ForWord than students with less profound language/speech and reading problems? In addition to evaluating the overall effect of the intervention, this design allowed us to test how well the major theory behind the program generalizes to an "at-risk" student sample and enabled us to explore for whom the intervention may be most helpful.

Fast ForWord effects were assessed on two outcomes. The program is intended primarily to improve the skills that are key predictors of reading success: phonemic awareness;

phonological awareness; and language skills. Our study tested to what extent the training helped students improve their language skills, but also tested whether the intervention helped students improve their reading comprehension skills, as measured by the Baltimore City's CTBS/5 Terra Nova testing program.

As suggested by the prior review of the Fast ForWord research base, the strongest connection between the theory behind the program and its potential promise seems to be for children experiencing speech and language difficulties in the classroom. Further, although there is the potential for Fast ForWord Language to improve reading skills, the theory would seem to suggest that students, in general, should achieve more direct and stronger improvements in language outcomes than reading outcomes, at least in the short term. The laboratory evidence supporting Fast ForWord and the program's theory of action suggest that targeting and remediating language skills, especially for students who are struggling with language and general literacy, should lead to gains in the more proximal outcome of the intervention, language skills, and to eventual gains in the more distal outcome of reading comprehension. Consistent with the program theory and the evidence supporting it, we hypothesized that all students participating in the program would achieve greater academic gains in language than reading comprehension and that students with more severe pre-program language and literacy difficulties would experience the greatest overall benefits from the intervention.

Finally, because the intervention was implemented in school-based settings rather than more tightly controlled clinical or laboratory settings, which were characteristic of earlier studies of Fast ForWord by researchers such as Merzenich et al. (1996) and Tallal et al. (1996), we suspected that compliance with the demanding treatment regimen would be relatively variable across students.[2] This issue raises the importance of collecting data that can help differentiate between the treatment that participants received as a result of random assignment and the treatment that they received in reality. To understand the intention-to-treat (ITT) effect and the effect of treatment on the treated, we applied two classes of analytical methods to the resulting data.

The effects estimated through the ITT analysis provide educators and policymakers an indication of the impacts that can be expected in typical school settings beyond the current experiment in which students and schools often do not fully comply with program assignments. To complement the ITT analysis, we used an instrumental variables approach to estimate the impact that one can expect when students receive the full treatment. In a randomized experiment with full treatment compliance, analyzing this result is simply a matter of comparing the treatment and control groups' outcome measures as in an ITT analysis. In most field-based studies—including the current one—assignment to treatment and control conditions is random but treatment compliance is not perfect (Borman, 2003). In these cases, the instrumental variables approach described by Angrist, Imbens, and Rubins (1996) can provide a useful analytical framework for estimating the causal effects of the treatment on the treated.

Method

Sample

The Fast ForWord Language intervention was offered to students from 8 schools within the Baltimore City Public School System (BCPSS): 2 elementary schools; 3 middle schools; and 3 elementary/middle schools. The student sample included a total of 415 children from two grade-specific cohorts: a group of 141 second grade students; and 274 seventh grade students.

---

[2] The hypothesized greater variability in school-based settings than laboratory settings is also consistent with the

These grade levels were selected to represent the elementary and middle school levels.

In an attempt to define a sample that was similar to Fast ForWord's typical school-based target population, second and seventh grade students from the eight schools were deemed eligible for the intervention if they had scored below national norms (i.e., the 50^{th} percentile) on the Total Reading outcome for the district-administered Comprehensive Test of Basic Skills, Fifth Edition (CTBS/5) during the spring of 2000. The 415 students selected to participate in the study had a spring 2000 CTBS/5 Total Reading national percentile score of 16.11, which was far below the national average. In addition to below-average CTBS/5 Total Reading outcomes, these students also tended to have below-average outcomes on language skills. The 415 students in the sample had a mean spring 2000 CTBS/5 Language Expression national percentile score of 22.13.

Measures

Student Achievement. Pretest data on experimental and control students' Reading Comprehension and Language skills were obtained from the spring 2001 administration of the CTBS/5, form B, by the BCPSS. These data were collected just prior to the start of the Fast ForWord program, which began on April 17, 2001. At the end of the week of June 11-15, 2001, upon completion of the Fast ForWord program, experimental and control students' Reading Comprehension and Language skills were retested using form A of the CTBS/5. Thus, the April and June tests represented, respectively, the pretest and posttest.

The CTBS/5, published by CTB Macmillan/McGraw-Hill, is a widely used and highly regarded achievement test. Prior analyses of national norming sample data indicate high item and scale reliabilities, absence of ceiling and floor effects, absence of obvious cultural biases,

---

findings of Rouse & Krueger (2003), who found notable variation among students with respect to their compliance and a low error of measurement. The test items from the CTBS/5 reflect the educational objectives that are commonly found in state and district curriculum guides and in major textbooks, basal series, and instructional programs.

The CTBS/5 is vertically equated and provides outcomes in the scale score metric. We present student pretest data in Table 2 as scale scores. However, for the ordinary least squares regression analyses and instrumental variables analyses, we converted the pretest and posttest scale scores to standardized scores. We performed this conversion separately by grade cohort, such that both second and seventh grade cohorts have pretest and posttest scores with means of 0 and standard deviations of 1. We standardized these scores to help provide a more direct interpretation of the results.

Student Background Data. In addition to test scores, we collected demographic data on the students based on information abstracted from the BCPSS' Pupil Information File (PIF). Data abstracted from the district-compiled PIF included: gender, race/ethnicity, federal free-lunch program eligibility status, and 1999-2000 school year attendance rate. These data were used to describe the characteristics of the sample and to confirm that the process of random assignment elicited treatment and control groups that were from statistically similar backgrounds.

Language and Speech Skills Observational Survey. Prior to random assignment and the start of the Fast ForWord intervention, the classroom teacher who was most familiar with each of the 415 eligible children was asked to complete a 20-item questionnaire regarding the child's language and reading skills. This survey, which was developed by Scientific Learning Corporation, included instructions to teachers stating that the questions "help to indicate whether

---

with the Fast ForWord training schedule when the program was implemented in an urban school district.

an individual has difficulty with learning language and reading skills…If you mark "sometimes" or "frequently" to many of the questions, the individual may be a candidate for a Scientific Learning training program." Though Scientific Leaning Corporation provides no evidence of the reliability or validity of the instrument, the fact that the company constructed this survey to determine whether a student is experiencing difficulty with language and reading and that it suggests using it to determine which students may be in need of the training program indicates some degree of content validity.[3]

Students were assigned a value of 1 when teachers indicated that the problem behavior occurred "rarely," a value of 2 when the teacher indicated that the behavior occurred "sometimes," and a score of 3 when the teacher indicated that the behavior occurred "frequently." The survey, thus, had a range of 20 to 60, with higher scores representing a greater prevalence of behaviors indicating difficulties with speech, auditory, language, and reading skills. The internal consistency reliability of the items in the scale was quite good, as suggested by the computed coefficient of reliability, Cronbach's $\alpha = .82$. The mean score for the sample was 34.60 with a standard deviation of 9.99. To facilitate interpretation of the survey scores, we employed a standardized score, with a mean of 0 and standard deviation of 1, in the analyses of Fast ForWord treatment effects.

Procedure

Random Assignment. The randomization process was conducted separately within each school and grade level with an equal selection probability of approximately 50% in all cases. As shown in Table 1, this combination of grade level by school clusters resulted in a total of 11

---

[3] Though the developer provided no empirical evidence, the survey score was consistently a statistically significant predictor of reading comprehension and language posttest outcomes across all regression models tabulated in Tables 3, 4, 6 and 7. Thus, in the current study, the instrument also exhibited good criterion-related validity.

randomization blocks or pools. Within each of the 11 grade level by school randomization blocks, approximately 50% of the students eligible for the intervention were assigned at random to the experimental group and 50% were assigned at random to the control group. This method of assignment was both fair and ethical, due to the fact that there were more eligible students than slots available in the program and that the intervention was, strictly speaking, of unproven educational value to the students. Summary data on the numbers of eligible students, the selection probabilities used, and the resulting experimental and control sample sizes are provided in Table 1.

Insert Table 1 about here

We conducted experimental-control comparisons of family background characteristics, pre-treatment language and speech skills survey data, and the CTBS/5 pretest scores to examine the extent to which the randomization process yielded statistically equivalent control and treatment groups. Table 2 shows the family background and baseline school participation and language skills of students randomized into the treatment and control conditions. The baseline characteristics of the total sample, as originally randomized, and the analytic samples, as used in the final analyses of treatment effects, are displayed for the second grade sample and the seventh grade sample. As the table shows, the participating students were predominantly African American and poor, as indicated by eligibility for the federal free-lunch program. The data displayed for the baseline samples revealed that randomization produced experimental and control groups that were roughly equivalent on all pre-intervention measures. Using a Bonferroni adjustment for multiple comparisons, our statistical comparisons of experimental-control differences for regular school year attendance rates, *t*-tests of baseline survey outcomes and achievement scores, and χ^2 analyses of the experimental-control differences on gender, ethnicity, and free lunch status revealed no differences at $p < .05$.

Listwise deletion of student cases with missing pretest-posttest data did not cause differential attrition rates by program condition for the second grade cohort, $\chi^2(1, N = 141) = 0.70$, $p = 0.40$, leaving 79% of the baseline sample of 71 treatment students and 73% of the 70 baseline controls for the impact analyses. Data attrition for the seventh grade sample also was statistically equivalent across treatment and control, $\chi^2(1, N = 274) = 1.83$, $p = 0.18$, leaving 62% of the 139 treatment students from the baseline sample and 70% of the 135 baseline controls. Furthermore, the pretest scores, regular school year attendance rates, baseline survey outcomes, and gender, ethnicity, and free lunch status of those treatment students who were retained for the analyses and of those control students who were retained for the analyses were statistically equivalent for both second and seventh graders. In these respects, the equivalence of the experimental and control samples produced through randomization was not undermined by pretest-to-posttest data attrition.

Finally, to assess the external validity of the study, we looked at whether treatment and control students lost from the sample due to data attrition were systematically different from students who remained in the sample in terms of their background characteristics, 1999-2000 school attendance, and teacher-appraised language skills. The results revealed one difference for second graders, and several differences between seventh grade students with complete and incomplete data. Among second graders, African American students were more likely to remain in the sample and white students were more likely to leave $\chi^2(1, N = 141) = 9.11$, $p < .01$, than expected assuming independence between race/ethnicity and sample membership. Among seventh graders, the average teacher survey score for seventh grade students with complete data was lower (*M* = 31.80, *SD* = 9.65) than that for students with incomplete data (*M* = 37.18, *SD* = 9.57). This difference was statistically significant, $t(274) = 4.40$, $p < .001$ (two-tailed), indicating that those who remained in the seventh grade analytical sample were more skilled at language than their counterparts who were omitted from the analytic sample due to incomplete data. Similarly, the seventh grade students with complete data attended school at a higher rate (*M* = 0.92, *SD* = .07) than students with incomplete data (*M* = 0.85, *SD* = 0.11) during the preceding school year, $t(274) = -5.43$, $p < .001$ (two-tailed). Seventh grade students with complete data also had higher language pretest scores (*M* = 626.96, *SD* = 42.70) than students with incomplete data (*M* = 612.79, *SD* = 37.94) , $t(246) = -2.51$, $p < .05$ (two-tailed). Finally, the seventh grade students with complete data had higher reading pretest scores (*M* = 640.76, *SD* = 28.90) than students with incomplete data (*M* = 619.58, *SD* = 39.45) , $t(245) = -3.96$, $p < .001$ (two-tailed).

Though the differences, most notably for the seventh grade sample, may compromise the external validity of our analysis to some extent, there is no conflict in this experiment between random assignment of treatment and missing at random. That is, among the complete data observations, those assigned to treatment have similar covariate distributions to those assigned control. As noted by Rubin (1976) and Little and Rubin (1987), the missing data process is *ignorable* if, conditional on treatment and fully-observed covariates, the data are *missing at random* (MAR).

Insert Table 2 about here

Implementation of the Intervention. After randomization, students selected for treatment were offered the Fast ForWord Language program in resource rooms within each school that were designed to accommodate the students, computers, and software. Treatment students attended Fast ForWord training as a targeted pullout program offered during the regular school day. In general, while those students assigned to participate were pulled out of their regular classroom for the Fast ForWord training, control students received non-literacy instruction or participated in special activities and classes, such as art and gym. In this way, Fast ForWord typically supplemented the regular classroom literacy instruction afforded treatment students.

We conducted three brief site visits, during the first, third, and fifth week of the program, at each of the 8 schools implementing Fast ForWord Language. Using a checklist based on information provided by the Scientific Learning Corporation in its Site Compliance Guide, the primary purpose of these visits was to ascertain the quality of the program implementation at each site.[4] Following the protocols established in the Site Compliance Guide, we ensured that the computers and training rooms complied with the requirements specified by Scientific Learning. By noting each site's compliance with the requirements, we were able to ascertain that each school had achieved a basic level of program implementation. Additional questions on our Fast ForWord Site Observation Form noted the level of overall support for implementation that the school personnel received from Scientific Learning and the district, particular problems with the implementation, and, if problems had occurred, what had been done to remedy them.

Prior to the start of the program, a thorough training session was provided by Scientific Learning for all teachers operating the programs at the 8 Baltimore schools. Each of the computer labs was well equipped with on-line computers containing adequate memory and

[4] We consulted the "Site Compliance Guide," dated May 25, 2001, which is available from Scientific Learning.

RAM. Each training room had adequate lighting and presented limited visual and auditory distractions. At each school, there were more than adequate numbers of headphones for all students and teachers. The headphones had padded headbands and ear cups and covered the child's entire ear to maximize hearing. Although each computer did not necessarily have a Y-adapter to facilitate listening in by the Fast ForWord trainer, there were multiple adapters that trainers used at each computer as the need arose. In two schools, Y-adapters were not necessary, as the computers had connections for more than one set of headphones. All teachers reported success in connecting to Scientific Learning servers to receive daily updated student portfolios.

At all 8 schools, a representative from Scientific Learning had visited the school to oversee the initial implementation. None of the teachers expressed anything negative about the level of the support they were receiving from the district or from Scientific Learning. Only a few implementation problems were reported during our site visits. At one elementary school, there was a temporary problem with the sound on six of the computers that was related to a software malfunction. At one of the middle schools, implementation started a week later than scheduled because there were problems with the Internet connection. The staff from Scientific Learning, in cooperation with the technical staff from the district, were eventually successful in getting the computers online. At another school, the Fast ForWord demonstration program had to be used during the first week of implementation. Also, due to a corrupted file on one of the computers, the trainer had to access the exercises for the students manually. Aside from these minor and temporary problems, the software and Internet connections functioned properly.

Fast ForWord Attendance, Compliance, and Completion Data. We had unusually rich student-specific data to assess the fidelity of implementation and compliance with the treatment regime. We obtained data from the Scientific Learning Corporation regarding the progress each student was making in the Fast ForWord program. The company measures progress in three ways: the overall number of training sessions attended by the student; the student's compliance with the suggested program schedule; and the student's exercise completion rate. Scientific Learning Corporation recommends that elementary students train on Fast ForWord Language for 100 minutes per day, 5 days a week, for a minimum of 20 days, and middle- and high-school students train for 90 minutes per day, 5 days per week, for a minimum of 20 days.

The Fast ForWord compliance measure provides a single value that indicates how closely a student is following the developer's recommended training schedule. It is calculated by dividing the total minutes a student has trained by the total minutes a student should have trained, as recommended by the developer. For example, a student who trains 100 minutes per day 4 days a week will have a compliance percentage of 80% and a student who trains 5 days a week but leaves the Fast ForWord Language classroom 30 minutes early every day will have a compliance rate of 70%. Thus, compliance provides an important estimate of how diligently the students and the school are following the suggested training schedule.

In addition, for each exercise, the completion percentage is provided. For exercises that have processed speech, a completion percentage of 80% or more indicates normal speech. As a summary measure of each student's mastery of the program content, the percentage complete is averaged across all of the exercises. According to data provided by the developer, the national average completion percentage, across nearly 30,000 students, was 70%. Taken together, these attendance, compliance, and completion data allowed us to measure the extent to which each child was exposed to the intervention in a manner consistent with the recommendations of the developer.

Results

Data analysis began with a screening of students' pretest and posttest data. This analysis revealed that some students had demonstrated erratic pretest-to-posttest changes. These achievement differences fell outside the bounds of what one might expect, given the reliability of the tests being used, and the rather short period of time between the pretests and posttests. In order to identify outliers for removal from the analytic sample, we regressed posttest scores on the key variables in the analyses (i.e., pretest score, teacher survey score, gender, race, and free lunch status), with the exception of the treatment and compliance variables. After obtaining the regression coefficients, we computed the Cook's Distance statistic, *D*, for each observation from the two analyses of language and reading comprehension outcomes. Cook's *D* expresses both distance from the mean and leverage exerted upon regression coefficients. A case is considered to be a highly influential outlier if the value of *D* exceeds $4/n$ (Cook, 1977).

The study participants who exceeded this threshold tended to be students with average to slightly above-average skills, as indicated by the teacher survey, who experienced precipitous pretest-to-posttest declines in their achievement scores. When the *D* exceeded $4/n$ for either of the two analyses, or for both analyses, we eliminated the observation from the analytic sample. This procedure resulted in the elimination of 11 second grade students, 6 from the experimental group and 5 from the control group, and 21 seventh grade students, 12 from the experimental group and 9 from the control group. The outlier analysis removed students from the sample that exhibited performances on the posttest that were not credible, given their pretest performances.

We considered trimming or Winsorizing these extreme values and then performing the analyses with the trimmed values for the outliers. Anecdotal evidence from testing, though, suggested that many of these students suffered precipitous pretest-posttest achievement losses because they did not take seriously the end-of-school-year posttest. Trimming the extreme values for these outliers would not have corrected this problem and the trimmed values would not be extremely meaningful measures of the students' achievement outcomes. For this reason, we present the analyses with the outliers removed and provide the analytical outcomes for the total sample, including the outliers, in the Appendix.

Our primary estimates of the Fast ForWord program impacts were derived from intention-to-treat (ITT) analyses. Regardless of the experimental students' actual compliance with the Fast ForWord schedule and exercise completion rate, these analyses compared all students assigned to the experimental condition—those who were intended to receive the treatment—to all those assigned to the control condition. In experiments with imperfect compliance with the intervention's regimen, such as this one, the ITT effect represents only the impact of *assignment* to the treatment. This ITT effect provides a reasonable estimate of the overall effects on achievement that can be expected in the field when schools implement a Fast ForWord program with the typical, less-than-perfect student attendance and participation.

Additional analyses specifically addressed program attendance, compliance, and completion and how program completion may have influenced students' achievement outcomes. These analyses utilized an instrumental variables approach for estimating the treatment effect for the treated (Angrist, et al., 1996). In this formulation, assignment to treatment was modeled as an instrument for actual participation. Random assignment to Fast ForWord did co-vary with actual participation in the program, but it was not correlated with the error term in the outcome equation because it was determined randomly. Under reasonable assumptions, the instrumental variables model yields a precise and consistent estimate of the effect of "treatment on the treated." In our analyses, the second-stage, or outcome, equation was represented by models that included a mediator variable measuring actual participation and compliance with the treatment: the student's completion rate.

Analyses of Intention-to-Treat Effects

Tables 3 and 4 provide summaries of a series of ordinary least-squares regression models that tested the hypothesized main effects and interaction effects of treatment assignment on achievement. Table 3 presents results for the second grade cohort and Table 4 summarizes the outcomes for the seventh grade students. These models included measures of student background as covariates, which helped improve the precision of the Fast ForWord impact estimates.

Model I regressed language and reading posttest scores on the following set of covariates: pretest; teacher-rated speech/auditory/language difficulties; grade level; gender; free-lunch status; African American; and a series of dummy codes representing fixed effects for each school. Model II included these covariates along with an indicator of treatment assignment (i.e., Fast ForWord or control). Models III and IV added interaction terms for, respectively, treatment by pretest and treatment by teacher-rated speech/auditory/language difficulties. Thus, Model II tested for the main effect of Fast ForWord treatment and assignment, Model III tested for the pretest-by-treatment interaction effect, and Model IV assessed the speech/auditory/language-by-treatment interaction effect.

In addition to these fixed-effects ordinary least squares regression models, we estimated preliminary multilevel or hierarchical linear models, which nested students within the school-level randomized blocks. In this random-effects conception, the school-specific blocks were modeled as level 2 clusters in the multilevel analysis, with the Fast ForWord treatment effect specified as randomly varying across blocks. In this sense, any heterogeneity in the treatment impact across schools was modeled as a random effect. In contrast, the fixed-effects approach, described by Schochet (2005), is often more realistic in evaluations of education interventions. The current study, like most education evaluations, included a small number of purposively-selected schools. In many such evaluations, it is untenable to assume that the study sites are truly representative of a broader, well-defined population of sites. Furthermore, as Schochet pointed out, inflating the standard errors to incorporate between-block effects will slant the study in favor of finding internally valid impact estimates that are not statistically significant, thereby providing less information to policymakers on potentially promising interventions.

In random-effects designs, though, study results can be generalized more broadly than in the fixed-effects designs. In addition, modeling the Fast ForWord impact as a random effect across the school sites helped us explore directly the extent to which the treatment effect randomly varied across contexts or whether it was relatively stable and homogeneous across schools. However, these benefits involve costs in terms of precision, because the variance formulas must be inflated to account for between-site random effects rather than fixed effects. Intuitively, in repeated sampling, different sets of sites would be selected for the evaluation, which could produce variability in the impact findings. Therefore, the variance expressions must account for the extent to which mean student outcomes may vary across sites. With only 4 schools for the second grade sample and 7 schools for the seventh grade sample, we had limited statistical power to explore these random effects across blocks. These multilevel models, though, revealed no statistically significant school-to-school variability for the treatment effects. With this empirical evidence, along with the practical problem of the limited number of school-level sites, our primary analyses focused on the fixed-effect outcomes for the sample of schools involved in the study.

In Table 3, the ordinary least squares regression models showed no statistically significant differences between second grade experimental and control students' language and reading comprehension outcomes and revealed no interaction effects. As indicated in the results for Model II, there were small but non-statistically significant differences for treatment assignment that favored the Fast ForWord group with respect to the language outcome, with an effect size of $d = 0.08$, but that favored the control group by $d = 0.07$ for reading comprehension. Models III and IV in Table 3 showed no interaction effects for, respectively, treatment by pretest and treatment by teacher survey.

The results for seventh grade students displayed in Table 4 revealed statistically significant main effects of treatment assignment for reading comprehension but no statistically or practically significant impacts on the language outcome. As shown in Model II, the effect of Fast ForWord assignment on reading comprehension outcomes was equivalent to an effect size of $d = 0.21$. Model III indicated no treatment by pretest interaction and the results for Model IV indicated no treatment by teacher survey interaction. We jointly tested the statistical significance of the treatment coefficients and interaction terms for all of the models that included treatment, treatment-by-pretest, and treatment-by-teacher-survey interactions. The joint *F*-tests revealed only one statistically significant outcome for Model IV reported in Table 4 for the reading comprehension outcome, $F\ (2, 166) = 3.13$, ;$p < .05$.

In addition, we estimated all regression models for the full second and seventh grade samples, which included the deleted outliers. The results of these analyses are presented in Tables 7 and 8, which are found in the Appendix. Predictably, these models, which included the outlier cases, explained less variability in the posttest outcomes. This difference was especially pronounced for the seventh grade sample, which included a larger number of outliers than the second grade sample. The results shown in Table 7 for the full sample of second graders revealed that the sign and magnitude of the treatment impacts were generally similar to those presented for the restricted sample in Table 3. The outcomes in Table 8 for the seventh graders tended to yield larger standard errors for the treatment coefficients. In addition, the treatment impacts were slightly smaller in magnitude. As a result, the statistically significant treatment effects for reading comprehension, which were noted in Table 4 for the restricted sample of seventh grade students, were not robust to the inclusion of the outliers.[5]

Insert Tables 3 and 4 about here

Fast ForWord Attendance, Compliance, and Completion Data

Two keys to the success of the Fast ForWord program are, of course, diligent attendance by the students in the program and successful completion of the program's exercises. Table 5 documents attendance, compliance, and completion data for the second and seventh grade treatment samples. The column labeled "Attended Fast ForWord Program" reports the percent of treatment students who attended at least one day of the program. These data show that turnout for the program was close to100%.

The next column in Table 5 documents the average number of days of training that students received. With an average of over 23 days of training for second grade students and over 20 days of training for seventh grade students, both cohorts exceeded the minimum standard

[5] Statistical power analyses showed that the second and seventh grade sample sizes were more than adequate to detect an effect equivalent to one quarter of one standard deviation or greater. With the covariates in our model accounting for a minimum of one third of the variability on the posttest, an alpha level of $p < .05$, and assuming an effect size of $d = 0.25$, power exceeded 0.80 with a total sample of $n = 100$, with 50 treatment students and 50 controls. This effect size standard we have chosen is relatively small in magnitude, but it is also one that has been

of 20 days recommended by the developer. The next columns present the compliance and completion data for Fast ForWord students. Compliance was reasonably high, as both second grade and seventh grade students trained for over 75% of their allotted minutes, on average. Finally, second graders, on average, completed fewer exercises, 62%, than seventh graders, 71%. Thus, in this case seventh graders met the standard of 70% cited by Fast ForWord but second graders did not.

In addition to the means and standard deviations reported for days of training, compliance, and completion, Table 5 indicates the number and percentage of students achieving the Fast ForWord standards for each of these outcomes. As suggested by the data presented in Table 5, the clear majority of students met or exceeded the Fast ForWord standards of 20 days of training, a compliance rate of over 70%, and a completion rate of over 70%. The one exception was for second grade students, of whom only 30% met the standard for the Fast ForWord completion rate. Therefore, in general, student attendance, completion, and compliance was sufficient to meet Fast ForWord standards. Most students attended the program, persevered for the required number of weekly minutes, and successfully completed the exercises. Despite generally high attendance and compliance, though, many second grade students experienced difficulty meeting the exercise completion standard of 70%.

Insert Table 5 about here

Causal Effects of Program Completion

---

generally deemed of practical educational importance (Slavin, 1990). It is also considerably smaller than the effect of approximately $d = 0.80$, which was claimed by the Fast ForWord developers (Rouse & Krueger, 2004).

Table 6 summarizes the outcomes from our two-stage least squares regression models. In these models, assignment to treatment was modeled as an instrument for actual participation, with participation measured by the completion rate. A key assumption of the instrumental variables analysis—the exclusion restriction—is that the instrument has no direct effect on the outcome within levels of program completion—the putative causal variable (Angrist et al., 1996). To meet this assumption, we recognized that the instrumental variables analysis must specify one and only one causal variable per instrument. Therefore, we focused on the measure of completion, which was the clearest indicator of students' commitment to both following and successfully completing the Fast ForWord Language treatment regimen. We employed the same set of covariates as that used in the previous regression models for the ITT effects.

Table 6 presents the results for second grade and seventh grade students for both language and reading comprehension outcomes. The Fast ForWord program completion coefficients estimated for the second grade cohort's language and reading comprehension did not reach conventional levels of statistical significance. Therefore, among second grade students assigned to Fast ForWord, program completion did not mediate the literacy effects as measured by the CTBS/5.

Insert Table 6 about here

The results for seventh graders indicated statistically significant effects of program completion on reading comprehension. The coefficient for the completion mediator variable suggests that a one standard deviation increase in the completion rate, or a 17 percentage point increase, was associated with an effect size of d = 0.11. Thus, beyond the ITT effects of treatment assignment for seventh graders, the instrumental variables analyses suggested a mediating effect of Fast ForWord participation, as measured by the program completion rate. The analyses tabulated in Table 9 in the Appendix, though, suggested that after inclusion of the outliers, this result did not hold. Similar to the case with the ITT analyses for seventh graders, the effect of the treatment on the treated was not robust to inclusion of the outliers.

Discussion

From the beginning, the implementation of the program and the level of support from within the school system and from the program developer, Scientific Learning Corporation, were exemplary. Site visits, observations of the training, communications with the Scientific Learning Corporation, and communications with the teachers and principals implementing Fast ForWord revealed a consistent level of commitment and support across district leadership, school-level leaders, the schools' teaching staffs, and representatives from Scientific Learning Corporation. Nevertheless, impact analyses of assignment to the Fast ForWord program revealed few encouraging signs of academic benefits approaching those claimed by the program's developers. In this way, the results raise several questions regarding the potential and the appropriateness of Fast ForWord for improving reading and language outcomes for non-clinical, "at-risk" student populations served in school-based settings.

There are several important explanations for this finding. First, and most importantly, although the majority of students who were served by the program met the developers' standards for Fast ForWord attendance, compliance, and completion, a sizeable proportion did not. These outcomes are consistent with those reported by Rouse and Krueger (2004), the authors of the only other randomized evaluation of Fast ForWord conducted independently of the Scientific Learning Corporation. These researchers also evaluated school-based implementations of the program within a large urban school district in the Northeast. Across two successive waves of implementation, the authors noted that between 67% and 76% of students attended the requisite 20 days of training and only 38% to 51% completed a sufficient proportion of exercises as suggested by the developers. Though we observed somewhat higher attendance and completion rates, this is likely to be primarily a function of the fact that students in our study had 8 weeks of potential training days rather than 7 weeks in the Rouse and Krueger (2004) study.

One of the most common difficulties at the schools was student scheduling for Fast ForWord training. This was especially the case at the middle schools. At one school, for instance, it took nearly two weeks before all regular classroom teachers fully understood the new schedule and consistently sent students to Fast ForWord at the appointed time. In addition, students were, at times, resistant to attend because they often had to miss physical education or other electives to come to the Fast ForWord training.

A final challenge for the Fast ForWord teachers was simply motivating the students to come to the resource room and to do the work. In 4 of the 8 schools, teachers reported that students were excited about the training initially, but that they soon complained of being bored by the repetitive nature of the activities. This pattern was most prevalent among the seventh grade students. At one middle school, some students had complained that the computer graphics were “babyish” and that their assignment to the program had made them feel “stupid.” The token economy rewards and consistent words of encouragement from teachers were important sources of external motivation that helped students remain fairly motivated and on task.

The lack of consistent effects of Fast ForWord Language for elementary and middle-school students is certainly policy-relevant in its own right, but this average effect of the intent to treat is best understood as an amalgam of treatment effects for those who completed the training with greater and lesser success. Our supplementary analyses, which examined the causal effects of participation, revealed that when the middle school teachers and students remained committed and more faithfully achieved the completion standards set by Scientific Learning Corporation, that the students exhibited statistically significant improvements in reading comprehension. Though evidence from this study and from the study conducted by Rouse and Krueger (2004) suggests that the demanding Fast ForWord training regimen can be difficult to schedule and implement in school-based settings, our results provide some evidence for seventh graders to suggest that when it is successfully carried out students' literacy outcomes can improve.

A second explanation for the results from our intent-to-treat analyses may be that Fast ForWord Language is designed to help students who have very specific and profound learning problems, which were relatively rare in this sample of students who were, more generally, at risk. The real strength of the Fast ForWord research base and the theory behind the program are in its application to children with speech/language/auditory difficulties. The program's potential to impact a more general population of educationally at-risk students is less established. Though the aptitude-by-treatment interaction effects that we tested revealed no reliable differences in program impacts across students with varying levels of pre-program reading achievement and language difficulties, the language-impaired students who experienced large effects in past studies by Merzenich et al., (1996) and Temple et al. (2003) represented only a very small subpopulation from this larger population of students at risk for poor literacy outcomes. If this interpretation holds true in other contexts, then the developers' ongoing efforts to market and disseminate the program widely to broad populations of students across the U.S. who are experiencing reading difficulties—rather than the far smaller population of students experiencing very specific language impairments—may be misdirected.

The consistent null effects on the language outcomes, combined with the fact that the one statistically significant intention-to-treat effect was for seventh grade reading comprehension, was also unexpected. The program theory stresses language achievement as a more proximal outcome of Fast ForWord than reading comprehension. However, the middle and high school variant of the Fast ForWord Language program used by the seventh grade students in our study, along with the regular classrooms that served them, stressed reading skills to a greater extent than the program and regular classroom instruction serving the second grade participants. In this way, the stronger effects of Fast ForWord on seventh graders' reading comprehension skills, relative to second grade students' reading outcomes, seem plausible. The lack of program impacts at either grade level for the language outcome, though, was perplexing.

The final explanation for our findings, though, may relate to the specific nature of the language and reading outcomes we measured. Rather than oral language competencies, which appear to be most directly affected by Fast ForWord Language, this study concerned whether the program produced effects on group-administered standardized achievement tests that are typically used by school systems for accountability purposes. It is possible that the treatment students in our sample realized some important improvements in their language skills that the CTBS/5 was not sensitive enough to measure. However, the expectation that Fast ForWord Language can improve students' outcomes on state assessments and standardized norm-referenced tests of various literacy outcomes, including comprehension, language, oral reading fluency, and vocabulary is supported by the results presented on the developers' website (see http://www.scilearn.com/results/main=home/rd). Indeed, leaders from the Baltimore City Public School System, who had input into the design of the study, were most interested in Fast ForWord's effects on assessments such as these that had important consequence for students and schools in the district. Improved student performance on high-stakes tests and progress toward school improvement as measured by these local forms of accountability were the outcomes desired by the participants in Baltimore, and, most likely, by most school leaders who purchase the intervention.

Conclusion

The Fast ForWord Language program did not, in general, help students in these 8 schools improve their language and reading comprehension outcomes. This result raises questions concerning the appropriateness of Fast ForWord Language training for school-based populations of educationally at-risk students, especially if the key outcome of interest is improvement on district- or state-administered standardized tests used for accountability purposes.

Our results also raise questions regarding the viability of scheduling and implementing the demanding training schedule of 90-100 minutes per day, five times per week in an urban school setting. In most respects, this was an exemplary implementation of a school-based educational program. Yet, second grade students, in particular, had trouble meeting the developer's program completion standards and some seventh graders had difficulty attending the supplemental program at the prescribed time. As the instrumental variables analyses suggest, program completion mediates the effect of treatment assignment. This outcome may suggest that only a high level of day-to-day perseverance by teachers and students is likely to pay more consistent dividends in terms of improved literacy outcomes.

Finally, though those students assigned to Fast ForWord Language did not realize consistent achievement advantages over controls in the short term, we do not know whether these students will realize gains in the longer term. We also do not know what, if any, other skills Fast ForWord students may have improved. Are there more highly specific subgroups of students for which the program is especially well suited? Is it possible for Baltimore schools, and other urban schools in general, to achieve more consistent student participation and completion in the program and would such increases produce more robust impacts? Despite these questions, though, the results from this experimental study suggest that an implementation of Fast ForWord with a sample of educationally at-risk students from the urban school setting of Baltimore did not produce short-term achievement advantages on district-administered standardized tests that are used for accountability purposes. Combined with similar evidence from the randomized trial conducted by Rouse and Krueger (2004) within a separate sample of schools from the Northeast, this study reveals important limitations concerning the effectiveness of Fast ForWord as an intervention for turning around the reading outcomes of struggling readers from high-poverty, urban schools.

References

Adams, M.J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Angrist, J.D., Imbens, G.W., & Rubins, D.B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association, 91*, 444-472.

Begley, S. (2000). Rewiring your gray matter. *Newsweek*, January, 63-65.

Benasich, A.A., Thomas, J.J., Choudhury, N., & Leppanen, P.H.T. (2002). The importance of rapid auditory processing abilities to early language development: Evidence from converging methodologies. *Developmental Psychobiology, 40*, 278–292.

Borman, G.D. (2003). Experiments for educational evaluation and improvement. *Peabody Journal of Education, 77*(4), 7-27.

Calfee, R.C., Lindamood, P., & Lindamood, C. (1973). Acoustic-phonetic skills and reading: Kindergarten through twelfth grade. *Journal of Educational Psychology, 64*, 293-298.

Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics, 19,* 15-18.

Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., et al. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements.* San Francisco, CA: Jossey-Bass.

Fitch, R.H., & Tallal, P. (2003). Neural mechanisms of language-based learning impairments: Insights from human populations and animal models. *Behavioral and Cognitive Neuroscience Reviews, 2,* 155-178.

Ehri, L.C., & Wilce, L.S. (1980). The influence of orthography on readers' conceptualization of the phonemic structure of words. *Applied Psycholinguistics, 1*, 371-385.

Ehri, L.C., & Wilce, L.S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly 20,* 163-179.

Gillam, R.B. (1999). Computer-assisted language intervention using Fast ForWord: Theoretical and empirical considerations for clinical decision-making. *Language, Speech, and Hearing Services in Schools*, 30, 363-370.

Grigg, W.S., Daane, M.C., Jin, Y., & Campbell, J.R. (2003). *The nation's report card: Reading 2002* (NCES 2003–521).Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: John Wiley.

MacDonald, G.W., & Cornwall, A.(1995).The relationship between phonological awareness and reading and spelling achievement eleven years later. *Journal of Learning Disabilities, 28,* 523-527.

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

Perfetti, C.A., Beck, I, Bell, L., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33,* 283-319.

Reed, M.A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology*, *48,* 270-292.

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.

Rubin, D.B. (1976). Inference and missing data. *Biometrika, 63,* 581-592.

Schochet, P.Z. (2005). *Statistical power for random assignment evaluations of education programs.* Washington, DC: Mathematica Policy Research.

Shankweiler, D. P., Crain, S., Katz, L., Fowler, A. E., Liberman, A.M., Brady, S. Thornton, R., Lundquist, E., Dreyer, L., Fletcher, J., Stuebing, K.K., Shaywitz, S.E., & Shaywitz, B.A. (1995). Cognitive profiles of reading-disabled children: Comparison of language skills in phonology, morphology, and syntax. *Psychological Science, 6*, 149-156.

Slavin, R.E. (1990). IBM's Writing to Read: Is it right for reading? *Phi Delta Kappan ,72*(3), 214-216.

Tallal, P., Miller, S.L., Bedi, G., Byma, G., Wang, X., Nagarajan, S.S., Schreiner, C., Jenkins, W.M., & Merzenich, M.M.(1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Scienc*e, *271,* 81-84.

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language, 9,* 182-198.

Temple, E., Deutsch, G., Poldrack, R., Miller, S., Tallal, P., Merzenich, M., & Gabrieli, J., (2003). Neural deficits in children with dyslexia ameliorated by behavioral remediation: Evidence from functional MRI. *Proceedings from the National Academy of Sciences, 100*, 2860-2865.

Wagner, R.K., & Torgesen, J.K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*, 817-830.

Author Note

We thank Dr. Patty Abernethy, the former Southeast Area Executive Officer of the Baltimore City Public School System, for her support and for making this research possible. We also acknowledge the kind assistance of Barbara Bell-Holman, and the principals and teachers at the eight Baltimore City Public Schools served by the pilot Fast ForWord program. Funding for the study was provided by the Abell Foundation of Baltimore, Maryland. Any opinions expressed are those of the authors and do not necessarily represent those of the Baltimore City Public School System or the Abell Foundation.

Table 1

Summary of Sampling Frame.

School	Second Grade		Seventh Grade					
	Eligible Students	Selected for Program	Eligible Students	Selected for Program	Selection Probability	Total Eligible	Total Selected for Program	Total Controls
A	66	33			0.50	66	33	33
B			39	20	0.51	39	20	19
C			54	27	0.50	54	27	27
D	28	14	32	16	0.50	60	30	30
E	25	13	33	17	0.52	58	30	28
F			20	10	0.50	20	10	10
G	22	11	33	17	0.52	58	30	28
H			63	32	0.51	63	32	31
Total	141	71	274	139	0.51	415	210	205

Note. Nat Am = Native American; Af Am = African American; Reading Comp. = Reading Comprehension.

Table 2

Comparison of Background Data for Students in Experimental and Control Groups

		Family Background							School Participation & Skills		Literacy Achievement Pretests	
	n	Male %	Nat Am %	Asian %	Af Am %	White %	Latino %	Free lunch %	Attend *M* (*SD*)	Survey *M* (*SD*)	Language *M* (*SD*)	Reading Comp. *M* (*SD*)
Second Grade												
Total sample												
Experimental	71	52.1	0.0	0.0	88.7	11.3	0.0	73.2	0.93 (0.07)	35.87 (10.31)	592.24 (33.76)	587.83 (30.96)
Control	70	51.4	0.0	0.0	90.0	8.6	1.4	74.3	0.93 (0.06)	37.01 (9.40)	593.11 (31.00)	588.27 (25.48)
Analytic sample												
Experimental	56	51.8	0.0	0.0	92.9	7.1	0.0	78.6	0.93 (0.06)	35.52 (10.55)	592.71 (31.98)	588.14 (27.12)
Control	51	54.9	0.0	0.0	94.1	5.9	0.0	74.5	0.93 (0.06)	35.90 (8.18)	595.51 (29.33)	587.84 (22.88)
Seventh Grade												
Total sample												
Experimental	139	50.4	0.7	0.7	66.9	30.2	1.4	79.1	0.90 (0.09)	33.74 (9.91)	623.42 (41.85)	635.18 (33.75)
Control	135	48.2	1.5	1.5	68.2	27.4	1.5	83.0	0.90 (0.09)	33.56 (10.00)	622.90 (42.06)	635.10 (33.00)
Analytic sample												
Experimental	86	44.2	0.0	1.2	64.0	33.7	1.2	73.3	0.93 (0.07)	32.46 (10.24)	628.23 (42.70)	643.29 (25.70)
Control	94	45.7	2.1	1.1	66.0	29.8	1.1	84.0	0.92 (0.07)	31.20 (9.09)	625.80 (42.89)	638.45 (31.51)

Note. Nat Am = Native American; Af Am = African American; Reading Comp. = Reading Comprehension.

Table 3

Summary of regression analyses of the effects of assignment to Fast ForWord treatment on Language and Reading Comprehension outcomes for second grade students (*n* = 107).

	Model I		Model II		Model III		Model IV	
	Language	Reading	Language	Reading	Language	Reading	Language	Reading
Constant	0.077	0.407	0.036	0.442	0.066	0.434	0.032	0.459
	(0.401)	(0.412)	(0.411)	(0.422)	(0.413)	(0.427)	(0.414)	(0.425)
Pretest	0.393***	0.393***	0.395***	0.393***	0.479***	0.374*	0.396***	0.390***
	(0.093)	(0.091)	(0.093)	(0.092)	(0.130)	(0.144)	(0.094)	(0.092)
Teacher survey	-0.259**	-0.273**	-0.256**	-0.274**	-0.261**	-0.275**	-0.241	-0.335*
	(0.092)	(0.092)	(0.092)	(0.093)	(0.092)	(0.094)	(0.148)	(0.151)
Gender (male = 1)	-0.012	0.073	-0.009	0.070	-0.012	0.075	-0.011	0.081
	(0.170)	(0.171)	(0.171)	(0.172)	(0.171)	(0.175)	(0.173)	(0.174)
Free lunch eligible	0.273	-0.022	0.265	-0.015	0.258	-0.011	0.270	-0.033
	(0.217)	(0.223)	(0.219)	(0.224)	(0.219)	(0.227)	(0.223)	(0.228)
African American	-0.596	-0.506	-0.591	-0.511	-0.612	-0.508	-0.591	-0.511
	(0.372)	(0.379)	(0.374)	(0.381)	(0.375)	(0.384)	(0.376)	(0.383)
Treatment			0.079	-0.068	0.080	-0.068	0.079	-0.067
			(0.162)	(0.164)	(0.162)	(0.165)	(0.163)	(0.165)
Treatment x pretest					-0.152	0.030		
					(0.165)	(0.174)		
Treatment x teacher survey							-0.023	0.093
							(0.177)	(0.180)
School 2	0.589*	0.334	0.594*	0.338	0.588*	0.336	0.596*	0.334
	(0.253)	(0.254)	(0.254)	(0.255)	(0.254)	(0.257)	(0.256)	(0.256)
School 3	0.366	-0.129	0.366	-0.128	0.348	-0.131	0.372	-0.151
	(0.214)	(0.217)	(0.214)	(0.218)	(0.216)	(0.220)	(0.220)	(0.224)
School 4	0.702*	0.111	0.697*	0.115	0.678*	0.120	0.702*	0.100
	(0.266)	(0.274)	(0.267)	(0.275)	(0.268)	(0.278)	(0.270)	(0.278)
R^2	0.365	0.343	0.366	0.344	0.372	0.344	0.366	0.346

Note. ***p < .001; ** p < .01; * p < .05; *SE*s in parentheses.

Table 4

Summary of regression analyses of the effects of assignment to Fast ForWord treatment on Language and Reading Comprehension outcomes for seventh grade students (*n* = 180).

	Model I		Model II		Model III		Model IV	
	Language	Reading	Language	Reading	Language	Reading	Language	Reading
Constant	-0.003	0.014	-0.041	-0.120	-0.029	-0.120	-0.039	-0.129
	(0.216)	(0.185)	(0.230)	(0.194)	(0.230)	(0.195)	(0.230)	(0.194)
Pretest	0.222**	0.603***	0.222**	0.595***	0.279**	0.600***	0.220**	0.606***
	(0.072)	(0.062)	(0.072)	(0.062)	(0.091)	(0.073)	(0.073)	(0.062)
Teacher survey	-0.333***	-0.084	-0.335***	-0.094	-0.330***	-0.093	-0.346**	-0.015
	(0.071)	(0.061)	(0.072)	(0.060)	(0.072)	(0.060)	(0.099)	(0.083)
Gender (male = 1)	-0.399**	-0.233*	-0.397**	-0.227*	-0.392**	-0.226*	-0.396**	-0.231*
	(0.121)	(0.104)	(0.122)	(0.103)	(0.122)	(0.104)	(0.122)	(0.103)
Free lunch eligible	-0.106	-0.198	-0.097	-0.170	-0.102	-0.171	-0.099	-0.160
	(0.162)	(0.137)	(0.163)	(0.136)	(0.163)	(0.138)	(0.164)	(0.136)
African American	0.062	-0.108	0.066	-0.098	0.050	-0.099	0.064	-0.088
	(0.140)	(0.119)	(0.140)	(0.118)	(0.141)	(0.118)	(0.141)	(0.118)
Treatment			0.060	0.213*	0.058	0.213*	0.060	0.212*
			(0.120)	(0.102)	(0.120)	(0.102)	(0.120)	(0.101)
Treatment x pretest					-0.122	-0.015		
					(0.119)	(0.106)		
Treatment x teacher survey							0.021	-0.138
							(0.120)	(0.102)
School 2	-0.293	0.281	-0.288	0.295	-0.289	0.298	-0.288	0.298
	(0.198)	(0.170)	(0.199)	(0.169)	(0.199)	(0.170)	(0.199)	(0.168)
School 3	-0.036	-0.248	-0.042	-0.265	-0.024	-0.261	-0.042	-0.271
	(0.243)	(0.210)	(0.244)	(0.208)	(0.244)	(0.210)	(0.245)	(0.207)
School 4	0.203	0.558**	0.196	0.537**	0.206	0.538**	0.196	0.535**
	(0.226)	(0.195)	(0.226)	(0.194)	(0.227)	(0.194)	(0.227)	(0.193)
School 5	1.058***	0.696**	1.060***	0.712***	1.057***	0.715***	1.059***	0.714***
	(0.228)	(0.202)	(0.229)	(0.200)	(0.229)	(0.201)	(0.230)	(0.199)
School 6	0.571	0.616*	0.568	0.610*	0.564	0.614*	0.569	0.606*
	(0.300)	(0.258)	(0.301)	(0.256)	(0.301)	(0.258)	(0.302)	(0.255)
School 7	0.323	0.406*	0.321	0.402*	0.324	0.404*	0.320	0.414*
	(0.202)	(0.172)	(0.202)	(0.171)	(0.202)	(0.171)	(0.203)	(0.170)
R^2	0.421	0.574	0.422	0.585	0.426	0.585	0.422	0.590

Note. ***$p < .001$; ** $p < .01$; * $p < .05$; *SE*s in parentheses.

Table 5

Summary of Fast ForWord Attendance, Compliance, and Completion Data.

		Attended Fast ForWord Program	Days of Fast ForWord Training		Fast Forward Compliance Rate		Fast Forward Completion Rate	
				Met Standard		Met Standard		Met Standard
	N	*N* %	*M* (*SD*)	*N* %	*M* (*SD*)	*N* %	*M* (*SD*)	*N* %
Second Grade	56	56 100%	23.2 (3.8)	48 85.7%	77.5 (15.1)	40 71.4%	61.9 (15.7)	17 30.4%
Seventh Grade	86	85 98.8%	20.3 (5.9)	57 66.3%	76.4 (16.1)	61 70.9%	71.1 (16.7)	62 72.1%

Table 6

Summary of two-stage least squares regression estimates of the effects of the Fast ForWord completion rate on Language and Reading Comprehension outcomes for second grade (*n* = 107).and seventh grade students (*n* = 180).

	Second Grade Students		Seventh Grade Students	
	Language	Reading	Language	Reading
Constant	0.080	0.406	-0.017	-0.037
	(0.402)	(0.413)	(0.218)	(0.185)
Pretest	0.391***	0.394***	0.220**	0.589***
	(0.093)	(0.092)	(0.072)	(0.062)
Teacher survey	-0.257**	-0.274**	-0.336***	-0.096
	(0.092)	(0.093)	(0.072)	(0.060)
Gender (male = 1)	-0.011	0.072	-0.397**	-0.226*
	(0.171)	(0.172)	(0.121)	(0.103)
Free lunch eligible	0.267	-0.018	-0.093	-0.153
	(0.218)	(0.224)	(0.164)	(0.138)
African American	-0.591	-0.511	0.065	-0.101
	(0.374)	(0.381)	(0.140)	(0.118)
Completion	0.042	-0.036	0.032	0.113*
	(0.086)	(0.088)	(0.063)	(0.054)
School 2	0.591*	0.338	-0.284	0.308
	(0.254)	(0.255)	(0.199)	(0.169)
School 3	0.363	-0.126	-0.039	-0.255
	(0.214)	(0.218)	(0.243)	(0.208)
School 4	0.691*	0.120	0.203	0.563**
	(0.268)	(0.275)	(0.226)	(0.193)
School 5			1.060***	0.717***
			(0.229)	(0.200)
School 6			0.575	0.633*
			(0.301)	(0.256)
School 7			0.319	0.394*
			(0.202)	(0.171)
R^2	0.366	0.346	0.423	0.585

Note. ***p < .001; ** p < .01; * p < .05; *SE*s in parentheses.

Appendix:

Outcomes for Full Student Samples, Including Outliers

Table 7

Summary of regression analyses of the effects of assignment to Fast ForWord treatment on Language and Reading Comprehension outcomes for second grade students (*n* = 118).

	Model I		Model II		Model III		Model IV	
	Language	Reading	Language	Reading	Language	Reading	Language	Reading
Constant	0.329	-0.101	0.315	-0.045	0.347	-0.026	0.344	-0.013
	(0.343)	(0.372)	(0.352)	(0.381)	(0.352)	(0.384)	(0.353)	(0.381)
Pretest	0.368***	0.336***	0.369***	0.335***	0.497***	0.401**	0.365***	0.325**
	(0.088)	(0.091)	(0.088)	(0.091)	(0.127)	(0.147)	(0.088)	(0.091)
Teacher survey	-0.191*	-0.241*	-0.190*	-0.245*	-0.190*	-0.239*	-0.309*	-0.386*
	(0.089)	(0.093)	(0.090)	(0.094)	(0.089)	(0.094)	(0.136)	(0.146)
Gender (male = 1)	-0.045	0.178	-0.045	0.177	-0.049	0.164	-0.030	0.196
	(0.157)	(0.171)	(0.158)	(0.171)	(0.157)	(0.173)	(0.158)	(0.172)
Free lunch eligible	0.087	-0.038	0.086	-0.033	0.092	-0.041	0.064	-0.058
	(0.199)	(0.216)	(0.200)	(0.216)	(0.199)	(0.217)	(0.200)	(0.217)
African American	-0.628*	-0.050	-0.628*	-0.050	-0.666*	-0.057	-0.629*	-0.051
	(0.295)	(0.320)	(0.296)	(0.321)	(0.296)	(0.322)	(0.296)	(0.320)
Treatment			0.028	-0.115	0.030	-0.114	0.025	-0.118
			(0.154)	(0.166)	(0.153)	(0.167)	(0.154)	(0.166)
Treatment x pretest					-0.218	-0.100		
					(0.157)	(0.176)		
Treatment x teacher survey							0.189	0.220
							(0.162)	(0.176)
School 2	0.182	0.315	0.182	0.320	0.174	0.318	0.178	0.318
	(0.235)	(0.253)	(0.236)	(0.253)	(0.235)	(0.254)	(0.236)	(0.253)
School 3	0.314	0.068	0.315	0.067	0.298	0.078	0.269	0.013
	(0.210)	(0.229)	(0.211)	(0.229)	(0.211)	(0.231)	(0.214)	(0.233)
School 4	0.582*	0.047	0.581*	0.050	0.578*	0.042	0.561*	0.031
	(0.251)	(0.275)	(0.253)	(0.276)	(0.252)	(0.277)	(0.253)	(0.276)
R^2	0.363	0.249	0.363	0.252	0.375	0.254	0.371	0.263

Note. ***p < .001; ** p < .01; * p < .05; *SE*s in parentheses.

Table 8

Summary of regression analyses of the effects of assignment to Fast ForWord treatment on Language and Reading Comprehension outcomes for seventh grade students (n = 201).

	Model I		Model II		Model III		Model IV	
	Language	Reading	Language	Reading	Language	Reading	Language	Reading
Constant	0.112	0.233	0.227	0.114	0.226	0.116	0.232	0.116
	(0.230)	(0.215)	(0.245)	(0.228)	(0.246)	(0.228)	(0.245)	(0.229)
Pretest	0.150*	0.428***	0.147*	0.427***	0.143	0.480***	0.145*	0.425***
	(0.072)	(0.068)	(0.072)	(0.068)	(0.096)	(0.088)	(0.072)	(0.068)
Teacher survey	-0.280***	-0.153*	-0.279***	-0.156*	-0.279***	-0.154*	-0.329**	-0.184*
	(0.073)	(0.068)	(0.073)	(0.068)	(0.073)	(0.068)	(0.097)	(0.091)
Gender (male = 1)	-0.328*	-0.221	-0.333**	-0.216	-0.333**	-0.207	-0.332**	-0.216
	(0.126)	(0.118)	(0.125)	(0.118)	(0.126)	(0.118)	(0.126)	(0.118)
Free lunch eligible	-0.235	-0.365*	-0.265	-0.335*	-0.265	-0.349*	-0.268	-0.336*
	(0.172)	(0.160)	(0.173)	(0.161)	(0.174)	(0.162)	(0.174)	(0.161)
African American	0.109	0.030	0.102	0.037	0.103	0.034	0.100	0.036
	(0.141)	(0.132)	(0.140)	(0.131)	(0.141)	(0.131)	(0.141)	(0.132)
Treatment			-0.169	0.176	-0.169	0.175	-0.169	0.176
			(0.123)	(0.115)	(0.124)	(0.116)	(0.123)	(0.116)
Treatment x pretest					0.008	-0.110		
					(0.123)	(0.116)		
Treatment x teacher survey							0.097	0.054
							(0.123)	(0.115)
School 2	-0.323	0.038	-0.331	0.046	-0.331	0.061	-0.331	0.046
	(0.204)	(0.191)	(0.204)	(0.191)	(0.204)	(0.192)	(0.204)	(0.191)
School 3	-0.002	-0.401	0.005	-0.407	0.005	-0.392	-0.001	-0.410
	(0.249)	(0.236)	(0.249)	(0.235)	(0.249)	(0.236)	(0.249)	(0.236)
School 4	0.233	0.363	0.242	0.355	0.241	0.363	0.242	0.356
	(0.237)	(0.225)	(0.237)	(0.224)	(0.237)	(0.224)	(0.237)	(0.225)
School 5	0.867***	0.510*	0.855***	0.526*	0.855***	0.526*	0.853***	0.526*
	(0.236)	(0.228)	(0.235)	(0.227)	(0.236)	(0.227)	(0.236)	(0.227)
School 6	0.382	0.370	0.375	0.378	0.376	0.395	0.376	0.379
	(0.310)	(0.292)	(0.309)	(0.291)	(0.310)	(0.292)	(0.309)	(0.292)
School 7	0.211	0.193	0.211	0.194	0.211	0.203	0.200	0.188
	(0.207)	(0.194)	(0.207)	(0.193)	(0.207)	(0.193)	(0.207)	(0.194)
R^2	0.292	0.378	0.299	0.386	0.299	0.388	0.302	0.388

Note. ***$p < .001$; ** $p < .01$; * $p < .05$; *SE*s in parentheses.

Table 9

Summary of two-stage least squares regression estimates of the effects of the Fast ForWord completion rate on Language and Reading Comprehension outcomes for second grade (*n* = 118).and seventh grade students (*n* = 201).

	Second Grade Students		Seventh Grade Students	
	Language	Reading	Language	Reading
Constant	0.331	-0.110	0.157	0.186
	(0.345)	(0.374)	(0.233)	(0.216)
Pretest	0.367***	0.338***	0.150*	0.422***
	(0.088)	(0.092)	(0.072)	(0.068)
Teacher survey	-0.190*	-0.248*	-0.277**	-0.158*
	(0.090)	(0.094)	(0.073)	(0.068)
Gender (male = 1)	-0.046	0.180	-0.334**	-0.215
	(0.158)	(0.172)	(0.126)	(0.117)
Free lunch eligible	0.086	-0.034	-0.278	-0.321*
	(0.200)	(0.217)	(0.175)	(0.162)
African American	-0.628*	-0.050	0.105	0.034
	(0.296)	(0.322)	(0.141)	(0.131)
Completion	0.015	-0.062	-0.090	0.093
	(0.082)	(0.090)	(0.065)	(0.061)
School 2	0.180	0.325	-0.341	0.055
	(0.236)	(0.254)	(0.204)	(0.191)
School 3	0.313	0.072	0.003	-0.403
	(0.212)	(0.230)	(0.249)	(0.235)
School 4	0.578*	0.063	0.225	0.374
	(0.253)	(0.278)	(0.237)	(0.224)
School 5			0.856***	0.528*
			(0.236)	(0.227)
School 6			0.362	0.394
			(0.310)	(0.291)
School 7			0.219	0.186
			(0.207)	(0.193)
R^2	0.363	0.249	0.297	0.387

Note. ****p* < .001; ** *p* < .01; * *p* < .05; *SE*s in parentheses.