2011

# Two Studies of TierII Literacy Development: Throwing Sixth Graders a Lifeline

A. W. Graves

L. Duesbery

Nicole Pyle
*Utah State University*

R. Brandon

A. McIntosh



Recommended Citation
Graves, A. W., Duesbery, L., Pyle, N., Brandon, R., & McIntosh, A. (2011). Two studies of tier II literacy development: Throwing sixth graders a lifeline. The Elementary School Journal, 111(4), 641-661.

# TWO STUDIES OF TIER II LITERACY DEVELOPMENT

## *Throwing Sixth Graders a Lifeline*

**ABSTRACT**

Two experimental studies at one urban middle school investigated the effects of the combination of Tier I and Tier II evidence-based reading instruction compared to Tier I alone on struggling sixth-grade readers ($N = 109$). All participants received free or reduced-price lunch, and 95% were considered English learners at some point in their school history. In both studies, Tier II intervention consisted of intensive instruction in word analysis, fluency building, comprehension, and vocabulary for 30 hours across 10 weeks. Results of both studies taken individually and combined indicated significant differences in favor of the intervention groups on oral reading fluency. The second study indicated significantly stronger performances for the intervention group on the Woodcock Reading Mastery Test—Revised (WRMT-R/NU) passage comprehension subtest. Tier II interventions and Response to Intervention (RTI) for older struggling readers are discussed related to educational implications and future research.

Anne W. Graves
Luke Duesbery
SAN DIEGO STATE UNIVERSITY

Nicole B. Pyle
UNIVERSITY OF TEXAS–AUSTIN

Regina R. Brandon
Angela S. McIntosh
SAN DIEGO STATE UNIVERSITY

**ABOUT** 30% of high school students fail to graduate, and in high-poverty neighborhoods the number rises to about 50% (National Governor's Association, 2009). Evidence indicates that challenges for students with reading difficulties become apparent very early in the educational process and tend to evolve into persistent lifelong struggles (Fletcher et al., 2002; Graves, 2010). Multiple studies, meta-analyses, and reviews of literature in the last 10 years have

THE ELEMENTARY SCHOOL JOURNAL VOLUME 111, NUMBER 4

provided a virtual blueprint for reading instruction in the early grades consisting of intensive instruction with progress monitoring in four areas: word analysis, fluency building, reading comprehension, and vocabulary enrichment (Blachman et al., 2004; Deno, 2003; Gersten, Baker, Haager, & Graves, 2005; Graves, Gersten, & Haager, 2004; Torgesen et al., 2001; Vellutino, Scanlon, Zhang, & Schatschneider, 2008).

In recent years researchers have investigated older struggling readers and found that they appear to respond to interventions and strategies that are remarkably similar to those espoused for young readers. However, comprehension and vocabulary instruction appear to yield stronger effect sizes in middle school than in elementary school, and the reverse is true for word analysis (Edmonds et al., 2009; Scammacca et al., 2007). Similar types of intensive instruction have been found to be effective in raising reading ability in studies of English learners (Gersten et al., 2008; Graves, Plasencia-Peinado, Deno, & Johnson, 2005; Gunn, Biglan, Smolkowski, & Ary, 2000; McIntosh, Graves, & Gersten, 2007; Purcell-Gates, Duke, & Martineau, 2007; Schleppegrell & Go, 2007).

Given the plethora of evidence for improving reading trajectories, in addition to the recent reauthorization of the Individuals with Disabilities Education Improvement Act (IDEIA, 2006) including recommendations for the use of Response to Intervention (RTI), schoolwide implementation of literacy development for struggling readers is critical (Fuchs & Fuchs, 2006). In the early grades, RTI has been shown to be effective in providing literacy development to young struggling readers (O'Connor, Harty, & Fulmer, 2005; Vaughn et al., 2009). RTI is typically defined as a three-tiered approach that includes Tiers I and II in general education; literacy development by general education teachers based on evidence-based approaches defines Tier I. If students do not progress as expected in Tier I, Tier II is implemented, typically consisting of small-group instruction with specialized and more intensive literacy development strategies and practice. Tier III is designed for students who did not respond to Tier II intervention and has typically included special education services (Coyne, Kame'enui, & Simmons, 2004; Fletcher, Coulter, Reschly, & Vaughn, 2004).

Unfortunately, little success has been found in applying this approach in upper elementary and secondary grades (Bradley & Danielson, 2004; Denton, Wexler, Vaughn, & Bryan, 2008; Vaughn et al., 2010), yet a recent meta-analysis for secondary literacy instruction indicates strong effect sizes for older students given intensive interventions in reading (Edmonds et al., 2009). These data indicate the possibilities of improved literacy for older struggling readers. Therefore, when students transition from elementary school to middle school, in theory an analogous RTI-type model should be implemented. Throwing these students a new lifeline at the beginning of middle school is critical for facilitating literacy development in the hopes of increasing success through the end of high school (see Fig. 1).

The purpose of this article is to report our investigation of the effects of Tier I and Tier II instruction on sixth-grade struggling readers. A secondary purpose is to extrapolate from the data recommendations for the development of an analogous RTI model for middle school.

## Study Background

At the very beginning of our development process—about one year before the studies reported here were conducted (see Fig. 1)—a focus group of three sixth-grade

| Year One:<br>Focus Group and Pilot Study | Year Two: Study 1 | Year Three: Study 2 |
|---|---|---|
| Development of Tier I and<br>Tier II Intervention | Tier I plus Tier II<br>Intervention vs.<br>Tier I Control | Replication of Tier I plus<br>Tier II Intervention vs.<br>Tier I Control<br>• improved instructor training<br>• better reading comprehension measures |
| | Experimental Replication of<br>Pilot Study | Experimental Replication of<br>Study 1 |

Figure 1. Development plan.

teachers (one special education resource teacher, one English/language arts (ELA) teacher, and one social studies teacher) and two university researchers developed Tier I- and Tier II-type instruction for struggling readers in the sixth grade. In the spring of the development year, a pilot study was conducted to investigate the effects of the newly developed Tier I and Tier II interventions and to assist in further development (Graves, Brandon, Duesbery, McIntosh, & Block, 2007).

To assist the middle school with its problem of persistent underachievement of a high percentage of entering sixth graders, in the context of a pilot study (Graves et al., 2007) we began the development of a "bundle" of Tier II evidence-based practices designed for sixth graders who were achieving at about the third-grade reading level or below. We developed a Tier II intervention with the goal of implementing it in small homogeneous groups of struggling readers in sixth grade. Given convincing supporting evidence, we agreed that curricula should entail word analysis including options for beginning decoding and more advanced structural decoding, oral reading fluency, and reading comprehension with vocabulary development.

In our pilot intervention, our Tier II model was implemented in the context of quasi-experimental research (Graves et al., 2007). The intervention group in the pilot study received the Tier II intervention based on the principles described above. The control was a group of students at the school who were demographically similar, but there was no random assignment in this pilot study. Results indicated significant differences between the intervention and the control groups with about a 10-word-

per-minute gain for intervention group on oral reading fluency. However, MAZE results in the pilot study indicated no differences between the groups.

The results of this pilot study led us to be encouraged by the decoding and fluency interventions we had developed. In Study 1 reported here, we replicated the pilot intervention, adding random assignment to conditions, more vocabulary instruction, and a vocabulary measure. In Study 2 reported here, we replicated Study 1, adding more rigorous preparation for instructors and an additional comprehension measure, the Woodcock Reading Mastery Test—Revised (WRMT-R/NU) passage comprehension subtest. The research question we sought to answer is as follows: Do students who are far below and below basic in sixth grade who receive Tier I plus Tier II intervention in word analysis, fluency, comprehension, and vocabulary perform significantly better than those who receive Tier I only or business-as-usual instruction on ORF (oral reading fluency), MAZE syntactic knowledge test, vocabulary measures, and passage comprehension measures?

## Method

### Sixth-Grade Participants

Study 1 and Study 2 were both conducted in a large, urban, inner-city middle school; 100% of the incoming sixth graders received free or reduced-price lunch and 95% of the students were considered English learners at some point in their school history. The students were from cultural and ethnic groups from countries all over the world. This population is approximately 51% Latino, 25% African American and Northern African, 22% Indo-Chinese (Vietnamese, Cambodian, Laotian), and 2% European Anglo, who were typically Russian immigrants and other atypical Anglo-American students growing up in poverty. Six students in Study 1 and five students in Study 2 were technically rated by the California English Learners' Diagnostic Test (CELDT) at intermediate fluency in English, meaning they were nearly fluent in spoken English. Newcomers to the United States were not included in our studies as they received special English language development and intervention in another school setting.

In Study 1 (originally $N = 66$; after attrition $N = 59$) and in Study 2 (originally $N = 60$; after attrition $N = 50$), all student participants were designated "far below basic" and "below basic" based on the California Language Arts Standards Test at the end of fifth grade. "Far below basic" is defined by the state of California as seriously below grade level with a high risk of retention; "below basic" is defined as below grade level with a risk of retention. In both studies students were randomly assigned to one of two groups: intervention and control with stratification by gender, English learner designation, and learning disability status. In Study 1, for the intervention group the instructional treatment took place during the elective period or as part of an ELA block so as not to interfere with content courses. In Study 2, the intervention occurred during the second hour of the ELA 2-hour block in sixth grade.

Treatment group students were placed into homogeneous groups using pretest data and input from teachers. Groups of three were instructed for 3 hours per week over a period of 10 weeks. The control students attended their elective classes or English language arts classes as usual in Study 1. In Study 2, the control group at-

tended their English language arts classes as usual. All students in both treatment and control groups experienced a regular school day with all subject areas taught.

Students with learning disabilities in both the intervention and the control group were fully included in what was described by the school as a coteaching model. In this particular school, coteaching meant that students were in general education classrooms all day with some inconsistent support from a resource teacher or an aide in the context of individualized or small-group work within the general education classroom. Observations of classrooms and reports from teachers indicated that the support did not involve intensive intervention; rather, it consisted of assistance with classwork and assignments.

### Tier I Description

In our pilot study (Graves et al., 2007) and in both of the studies reported in this article (see Fig. 1), the school committed resources each year toward teacher improvement in Tier I, including literacy development workshops by the school district, and sixth-grade students were all required to take a 2-hour ELA block. In Tier I, the school administration made regular observations of teachers. By the third year of the development work, teachers in Tier I were required to produce positive results for students who had not reached proficiency in literacy with the threat of being dismissed if the goal was not accomplished. In addition, Tier I teachers were encouraged to form guided reading groups and to tailor assignments to the current literacy levels of students. In Tier I, ELA teachers gave weekly writing and reading assignments with the goal of improving the literacy trajectory of low-performing sixth graders. All students in both groups in our study received the school's Tier I intervention.

### Materials and Structure for Tier II Intervention

Given the results of our pilot study and information from the literature on the design of small-group instruction (Gersten et al., 2008; Vaughn, Linan-Thompson, & Hickman, 2003; Vaughn et al., 2003), for our Tier II intervention we designated three students to one instructor (3:1). Further, the instruction consisted of three 1-hour sessions a week for at least 10 weeks. Though other RTI models have reported Tier II intervention with five or more students, our thinking was that sixth-grade students who were 3 to 6 years behind in their literacy development needed more individualized assistance in small groups of three. Given the 1-hour time frame for intervention, our Tier II intervention consisted of homogeneous groups of three students that each day of the intervention received 1 hour consisting of (*a*) 20 minutes of word analysis including beginning decoding or structural decoding for 2.5-level readers and higher, (*b*) 20 minutes of fluency development, and (*c*) 20 minutes of reading comprehension with vocabulary development.

After much discussion about the large numbers of culturally and linguistically diverse students living in poverty in this school population and the extreme difficulty that many sixth graders were still having with reading, before the pilot study the focus group concluded that the following curricula components best exemplified the existing research for Tier II instruction in literacy development: Corrective Reading (Engelmann et al., 1999) or REWARDS (Archer, Gleason, & Vachon, 2002) for word

analysis, Read Naturally (Ihnot, Matsoff, Gavin, & Hendrickson, 2001) for fluency building, and Daybook (Spandel, Nathan, & Robb, 2001) for comprehension and vocabulary. Results of our pilot study indicated relative strength for our intervention; therefore, for Study 1 we wished to examine similar components in an experimental study with random assignment. Then, as the results were favorable in Study 1, we also implemented these same components in Study 2 with an additional measure.

**Word analysis instruction.** As in our pilot study, student participants received instruction in one of two decoding programs depending on their reading level. Corrective Reading (Engelmann et al., 1999) was implemented when students read at grade levels 0 to 2.4. Reading Excellence: Word Attack and Rate Development Strategies (REWARDS) (Archer et al., 2002) was implemented when students read at grade level 2.5 or higher.

Developed in the late 1960s, Corrective Reading has a long history of anecdotally reported and research-based success (Lloyd, Cullinan, Heins, & Epstein, 1980). Data suggest the success of the decoding strand of this material on improved phonics and phonemic awareness skills. Corrective Reading spans three levels: A, B, and C. For the purposes of this study, students reading from pre-primer to reading level 2.4 received instruction in either Decoding A (grade level 0 to 1.2) or Decoding B1 (grade level 1.3 to 2.4). This curriculum provided an adapted introduction of letter sounds, spacing the short vowel introduction to allow mastery of the most frequently appearing sounds before introducing the less frequently appearing sounds (i.e., the order of short vowel introduction is a, i, o, u, and e and is spaced across 60 lessons). Carefully designed reading passages provided controlled practice based only on the sounds that had been introduced. Corrective Reading, Decoding A and B1 included progress monitoring, abundant active involvement, and writing. Students used word-attack skills that merged phonemic awareness and phonics, and engaged in group readings, individual readings, and exercises to practice these skills. Instructors were prepared to follow the script and engage students in each activity just as the Corrective Reading teaching manual recommended. Each day, instructors taught a lesson for about 20 minutes, and the next session picked up where the lesson left off the day before. In Study 1, we placed students in Corrective Reading based on their ORF scores from the pretest. If students read 0 to 20 words per minute (wpm), they were placed in Level A. If they read 20 to 40 wpm, they were placed in Corrective Reading B. In Study 2, we used the decoding placement test in Corrective Reading to place students in either Corrective Reading A or B.

Recall that our plan was to use Corrective Reading for the lowest decoders and REWARDS for those reading at level 2.5 or higher because students reading at the latter level needed structural word analysis work on multisyllabic words. The REWARDS program is written to provide intensive short-term practice for struggling readers from reading level 2.5 and up (Archer et al., 2002). Though the scientific research to support the effectiveness of REWARDS is limited, the authors of the program have validated its effectiveness (Archer, Gleason, & Vachon, 2003).

The REWARDS program is built upon evidence-based practices of intensive practice and strategy instruction for structural decoding. In REWARDS, the instructor models a strategy to segment multisyllabic words into word parts. Students read the word part by part, applying the interactive strategies to break the word into word parts. The student learns to use this strategy to decode words in lessons 1–12. These

skills are in preparation for implementing their use when reading longer passages in lessons 13–25. Students practice phonemic awareness and phonics skills during the preskill activities. Both directly and indirectly, vocabulary and comprehension are addressed in the lessons. Each lesson includes spelling practice as well. Teachers followed the script and engaged students in each activity just as the REWARDS teaching manual recommended. Each day, teachers taught a lesson for about 20 minutes, and the next session picked up where they left off the day before. In 30 hours of instruction, most instructors finished the book.

**Fluency development.** Encouraged by the results of our pilot study, we again used an adapted version of Read Naturally (Ihnot et al., 2001) for fluency building, which was based upon scientific evidence about the need for fluency development in struggling readers. In contrast to early elementary school studies of Read Naturally (Denton, Fletcher, Anthony, & Francis, 2006), our pilot study indicated that the adapted implementation of Read Naturally (20 minutes a day) produced a significant gain in fluency for sixth graders. Therefore, we implemented it again in Studies 1 and 2 to test its effectiveness.

Read Naturally provided leveled reading passages with progress monitoring built in and comprehension and vocabulary exercises for each passage. In both studies, Read Naturally was adapted to fit our 20-minute time frame. Using the ORF pretest scores, students were placed in leveled readers at their approximate instructional reading level, where students were able to demonstrate 90%–95% accuracy on oral reading of the material. At the beginning of the lesson each day, instructors taught the key vocabulary listed at the top of the page. While students in the group were writing their predictions, the instructor had the students each read a 1-minute timing, or cold timing. Students charted their own scores. Next, the instructor and students read the entire passage and answered the four comprehension questions together. Instructors then led a discussion on the answers to the comprehension questions, returning to the text to provide evidence for the answers. In the end, while students wrote a "retelling" in their own words of the passage, instructors conducted a second 1-minute timing for each student, known as the hot timing. Students graphed that score in the same column as the first timing, but this time using red ink to make progress over the cold reading score more apparent. Hence, in both studies progress monitoring was built into the context of the Read Naturally lessons, including individualized daily graphing of the number of words read.

**Reading comprehension instruction and vocabulary development.** In Study 1, we decided to reimplement the procedure in our pilot, plus we added a vocabulary measure to determine if we could have had undetected results. Daybook for Critical Reading and Writing (Daybook; Spandel et al., 2001) was already developed and easy for the instructors in our pilot study to use (Graves et al., 2007). We selected it because of its focus on evidence-based practices with targeted vocabulary and specific, explicit reading comprehension skills instruction. The Daybook also included two- to three-page selections from Newbery Award–winning literature and other award-winning books for youth. This text provided an opportunity for the instructor to read aloud to the students in the context of modeling the use of reading comprehension strategies. All aspects of this curriculum were consistent with existing evidence for increasing reading comprehension and vocabulary.

The Daybook assisted instructors in exposing students to grade-appropriate literature while focusing on academic language and reading comprehension develop-

ment. The other 20-minute interventions already included in the bundle focused heavily on word analysis and fluency. Daybook passages and lessons were designed to immerse students in a variety of literature genres and contained short lessons in reading comprehension, critical reading, and vocabulary development. Using literature selections and targeted strategy instruction, reading and writing skill development was the major goal of Daybook. Students engaged in a number of activities in one lesson while the instructor either read text to them as they followed along or students were asked to take turns reading. For example, lessons provided explicit details on the steps of a strategy such as questioning, summarizing, or locating the main idea. In a typical lesson, students are asked to listen while the instructor discusses the steps of the strategy. Next, the instructor reads a short passage to the students and models the implementation of the strategy. Finally, for the duration of the short passage the students and the instructor continue to implement the strategy together. The publishers of Daybook state that it is designed to develop an appreciation for literature from a variety of genres and to foster a development of improved reading comprehension and an appreciation of language. As with the other segments, the instructor would pick up where the class left off in the preceding instructional segment.

In both studies, when students encountered words they didn't know, using the recommendations of Beck, McKeown, and Kucan (2008), instructors typically defined the unknown word, showed a picture or acted out the meaning, discussed the uses or multiple meanings of the word, and set up the use of the word in sentences. Students kept a list of the vocabulary taught. At the beginning of each comprehension lesson, students reviewed the vocabulary words from the previous lesson. In Study 1, instructors received several hours of training on the use and implementation of the Daybook lessons.

In Study 2, instructors received the same training, and all of the materials were used once again as described above. However, instructors in Study 2 received additional vocabulary instruction as well. Instructors were taught more about tiered analysis of vocabulary words and the value of defining words carefully, providing visuals, and providing deep-level practice (Beck et al., 2008).

## Measures

In both studies, students were given pretests, and once a random assignment was complete, pretest data were used for screening students to determine appropriate small grouping. We administered three reading measures in common to our two studies: a test of oral reading fluency (ORF), a test of vocabulary knowledge (VOC), and a test of syntactic reading comprehension (MAZE). These three measures aligned well with the intervention components and aligned more broadly to the focus of language arts instruction in sixth grade. To obtain a better estimate of reading growth during Study 2, we added the WRMT-R/NU passage comprehension subtest (Woodcock, 2003).

**Test of oral reading fluency (ORF).** The ORF measures a student's ability to read words from a printed passage. This measure has been recommended consistently due to its strong correlation with other more complicated reading measures including reading comprehension (Greenwood, Tapia, Abbott, & Walton, 2003) and for its ease of administration (Hasbrouck & Tindal, 2006). During individual administra-

tion, the student is presented with a passage of approximately 250 words and asked to read out loud for 1 minute. The number of incorrect words is subtracted from the number of correct words to arrive at a final score yielding total correct words read per minute. Words read differently due to regional dialect were not considered errors.

During both Studies 1 and 2, the same six ORF passages were administered: three passages before the intervention as the pretest, and three after during the posttest. The median score from each group of three was used as the student's final score. Passages used in both studies were obtained from the EasyCBM Web site (EasyCBM, 2009). The passages are known to be of comparable difficulty (Alonzo & Tindal, 2009) and are ideally suited to measuring changes in reading fluency over the course of each study. The passages used had already been confirmed to be reliable with Winsteps 3.61.1 analytic software employing a one-parameter Rasch model (Alonzo & Tindal, 2009). While the participants were all in sixth grade, prior experience working at the school led us to select third-grade passages to better align with the actual performance levels of the students eligible for Tier II intervention. Grade-level estimates for passages were obtained with the Flesch-Kincaid and Lexile readability indices.

**Test of vocabulary (VOC).** Our vocabulary instrument, used in both Studies 1 and 2, was originally field tested with 70 items. During the field test about 200 sixth-grade students were asked to identify from three choices the word that most closely matched a prompt word in meaning. The three possible answer choices include one correct, one near, and one far response. Subsequent to the field test, analyses were used to select the 25 best-performing words for inclusion on the measure used in this study. Item response data were modeled with a one-parameter Rasch model in Winsteps, v3.48. Items selected each had a mean square outfit between .5 and 1.5, and a set of approximately functioning distracters. Pre- and posttest forms of the vocabulary measure included the same set of items; however, the order of items was randomized in each version to minimize order effect.

The vocabulary measure was group administered in general education classroom settings following written standardized administration protocols both before and after the intervention period. Students were given as much time as they needed to complete the 25-item measure. Vocabulary measures were presented on one side of a single sheet of paper. Students selected their response by bubbling in the answer on the test sheet. Students who finished quickly were asked to draw a picture on the back of the test sheet.

**Test of syntactic reading comprehension (MAZE).** The MAZE measures a student's ability to select from a list of appropriate words to complete sentences. Students were given a standardized instruction protocol and a brief three-sentence practice session, and were then given 180 seconds to read quietly from a single page of text. The first sentence was left intact, and after that each seventh word was replaced by three words from which the student had to select the correct choice. During silent reading, the administrator monitored the students to make sure they were reading and circling responses. After 3 minutes the students were told to stop reading. A reading comprehension score was derived from the total words correctly chosen, less the errors. In this study, two third-grade MAZE measures from AIMSweb were administered to each student, one before intervention and one after intervention. Alternate form reliability for the MAZE measures used has been documented in the range of $r = .81$ (Shinn, Deno, & Espin, 2000).

**Passage comprehension subtest of the Woodcock Reading Mastery Tests—Revised/NU (WRMT-R/NU).** The WRMT-R/NU is a diagnostic reading battery commonly used in schools for literacy screening. For Study 2, we selected the passage comprehension subtest because of its relatively common use in similar research studies. The use of this subtest will assist researchers later in across-study comparisons. The passage comprehension test was administered individually. Students were asked to first read short sentences, and later longer passages, each with a missing word they were asked to provide as they read silently. Forms G and H were used for the pre- and posttests, respectively. Raw scores ranged from 0 to 68. Internal consistency reliability for Forms G and H is $r = .83$; it is reasonable to consider these forms parallel given the close proximity of the scale scores between the two forms across the same normative sample (Woodcock, 2003).

## Instructors

In both studies, graduate students ($n = 6$ in Study 1; $n = 5$ in Study 2) who each had a minimum of 20 prerequisite units in special education coursework and who were seeking credentials in special education provided the Tier II intervention. For Study 1 only, a doctoral student helped organize the intervention at the school and also instructed one group. All 11 instructors were female. In Study 1, one instructor was African American and two were Latina; the other three were of Anglo American descent. In Study 2, two were Latina and three were of Anglo American descent. The sixth graders were grouped according to reading levels, with 11 groups in Study 1 and 10 groups in Study 2. Otherwise, each instructor had two groups of students across two different periods in the school day, though in Study 1 the project manager also instructed one group. Instructors were randomly assigned to teach groups of sixth graders and given intensive training for 1 day in Study 1 and for 2 days in Study 2. Additional individualized training was provided as needed throughout the intervention.

In Study 1, the training for instructors occurred in a 1-day workshop, introducing the materials to them and allowing them to practice. Instructors were able to confer during the entire 10 weeks with a research manager who from time to time would model instructional procedures or answer clarifying questions about the interventions. For each small group of three students, a folder was generated with materials for each student for each of the components of instruction. In Study 2, a 2-day training allowed more time for modeling, practicing, and questioning.

## Fidelity of Treatment

In both studies, instructor logs indicated that they upheld the time restrictions on the interventions by 91% (Study 1) and 95% (Study 2). Observations of instructors were conducted twice a month using a checklist of practices that were taught in the in-service preparation before the training. The checklist was composed based on the standard activities and structure-typical lessons in our materials. Fidelity of implementation of instructional components and practices was 89%, meaning that instructors taught the three interventions for 20 minutes each per day with a majority of the required components and practices intact. The results of our observations and checklist scores were combined with our anecdotal field notes and used to provide

Table 1. Study 1: Growth on Three Reading Measures

| | | Pretest | | Posttest | | | |
|---|---|---|---|---|---|---|---|
| | *n* | Mean | *SD* | Mean | *SD* | Change | *d* |
| ORF: | | | | | | | |
| Treatment | 31 | 88.7 | 31.7 | 106.8 | 31.5 | 18.1* | .57 |
| Control | 28 | 102.1 | 32.2 | 103.2 | 32.4 | 1.1 | .03 |
| Vocabulary: | | | | | | | |
| Treatment | 31 | 14.1 | 4.1 | 17.3 | 4.5 | 3.2 | .75 |
| Control | 26 | 15.6 | 4.5 | 19.3 | 4.6 | 3.7 | .81 |
| MAZE: | | | | | | | |
| Treatment | 28 | 5.8 | 6.6 | 9.7 | 5.8 | 3.9 | .62 |
| Control | 24 | 9.2 | 6.6 | 12.3 | 7.8 | 3.1 | .43 |

Note.—Cohen's *d* is modeled as (posttest-pretest)/pooled *SD*.
* $p < .05$ interaction of within-groups pre- to posttest by between-groups condition.

feedback to instructors. Throughout the intervention, observers provided coaching and answered questions to clarify and improve the implementation of the Tier II instruction as it was originally designed. At times, particularly near the beginning of the intervention phase, instructors taught more than 20 minutes in a particular area and then in turn taught less in the other areas. Our observational data indicate that by month 2, instructors were better able to stay with the recommended time frame. In Study 2, there was 94% adherence for all aspects of the intervention.

## Results

In both studies, pre- and posttest score results from the ORF, VOC, MAZE, and WRMT-R were analyzed using mixed-model, repeated-measure ANOVAs (RM-ANOVA). Group comparisons between students in the treatment and control groups were reported. Group comparisons between treatment intervention students with and without learning disabilities were also reported. Interaction effects were analyzed as well. The standard $p < .05$ was used to test for statistical significance. For each measure, we modeled the pre- and postscore as the repeated measures, and group membership (treatment vs. control) as the between-subjects factor. Effect sizes were calculated using partial $\eta^2$ coupled with Cohen's *d*.

### Study 1

By the end of Study 1, of the original 33 students in the treatment intervention, 31 remained; of the original 33 in control, 28 remained ($N = 59$). Of those remaining, CELDT testing indicated that five students were rated at intermediate fluency in English. All the others had tested out of CELDT or were considered fluent English speakers. Three were in the intervention group and two in the control group.

**Oral reading fluency rates.** The RM-ANOVA revealed that growth in mean ORF scores within the intervention group was significant and depended on receiving the intervention, $F(1, 57) = 56.1$, $p < .05$, partial $\eta^2 = .50$, $d = .57$. ORF rates grew for both the treatment and control groups (see Table 1). The treatment group ($n = 31$) grew from a mean of 88.7 to 106.8, a gain of about 18 wpm over the intervention period. In contrast, the control group ($n = 28$) grew by about one word per minute during the same period, from a mean of 102.1 to a mean of 103.2.

There were nine students with learning disabilities in Study 1, five in the intervention group and four in the control. Students in the intervention group with learning disabilities ($n = 5$) grew from a mean ORF of 56.6 ($SD = 24.8$) to 78.4 ($SD = 25.2$), a growth of about 22 wpm. English learners ($n = 3$) in the intervention group gained 20, 18, and 15 wpm; they did not distinguish themselves from their peers.

**Vocabulary.** RM-ANOVA indicated no differences between groups, $F(1, 55) = 0.17$, $p = .68$, partial $\eta^2 = .03$, on the vocabulary measure. The number of words correctly identified grew for both the treatment and control groups. The treatment group ($n = 31$) grew from a mean of 14.1 to a mean of 17.3 on the 25-item test. Similarly, the control group ($n = 26$) grew from a mean of 15.6 to 19.3. Students with learning disabilities ($n = 5$) and English learners ($n = 3$) in the treatment group did not grow in vocabulary knowledge more than control students.

**MAZE comprehension.** RM-ANOVA indicated no differences between groups on our MAZE measure, $F(1, 50) = 0.44$, $p = .51$, partial $\eta^2 = .09$. The number of correctly identified words in the MAZE probe grew for both the treatment and control groups. The treatment group ($n = 28$) grew from a mean of 5.8 to 9.7 net correct words per 180 seconds. The control group ($n = 24$) grew from a mean of 9.2 to a mean of 12.4. For students with learning disabilities ($n = 5$) in the intervention group, MAZE test scores rose from 3.75 ($SD = 2.5$) to 6.50 ($SD = 3.87$). English learners ($n = 3$) in the treatment group did not grow more than controls.

## Study 2

By the end of Study 2, all 30 students in the treatment group remained. Of the original 30 control students, 20 remained ($N = 50$). Of those remaining, CELDT testing indicated that four students were at intermediate proficiency in English. All the others in the study had tested out of CELDT and were considered fluent in English. Two were in the intervention group and two in the control group.

**Oral reading fluency.** The RM-ANOVA revealed that growth in mean ORF scores within the intervention group was significant and depended on receiving the intervention, $F(1, 48) = 90.1$, $p < .05$, partial $\eta^2 = .55$, $d = .66$. ORF rates grew for both the treatment and control groups (see Table 2). The treatment group ($n = 30$) grew from a mean of 88.3 to 109.9, a gain of about 21.6 wpm over the intervention period. In contrast, the control group ($n = 20$) grew by 0.2 wpm during the same period, from a mean of 103.2 to a mean of 103.4. English learners in the intervention group ($n = 2$) grew by 20 and 18 wpm. Students in the intervention group with learning disabilities ($n = 5$) grew from a mean of 65 wpm ($SD = 42.2$) to 78.6 wpm ($SD = 48.6$), a growth of about 14 wpm.

**Vocabulary.** RM-ANOVA indicated no differences between groups on the vocabulary measure, $F(1, 45) = 0.34$, $p = .56$, partial $\eta^2 = .007$. The number of words correctly identified grew for both the treatment and control groups. The treatment group ($n = 29$) grew from a mean of 14.2 to a mean of 17.8 on the 25-item test. The control group ($n = 18$) grew from a mean of 15.9 to 20.1. English learners ($n = 2$) and students with learning disabilities ($n = 5$) receiving the intervention did not grow more on vocabulary knowledge than students in the control group.

**MAZE comprehension.** RM-ANOVA indicated no differences between groups on the MAZE measure, $F(1, 42) = 0.08$, $p = .77$, partial $\eta^2 = .02$. The number of correctly identified words in the MAZE probe grew for both the treatment and

Table 2. Study 2: Growth on Four Reading Measures

| | | Pretest | | Posttest | | | |
|---|---|---|---|---|---|---|---|
| | *n* | Mean | *SD* | Mean | *SD* | Change | *d* |
| ORF: | | | | | | | |
| Treatment | 30 | 88.3 | 31.2 | 109.9 | 33.8 | 21.6* | .66 |
| Control | 20 | 103.2 | 27.9 | 103.4 | 28.2 | .2 | .01 |
| Vocabulary: | | | | | | | |
| Treatment | 29 | 14.2 | 3.5 | 17.8 | 4.2 | 3.6 | .93 |
| Control | 18 | 15.9 | 4.4 | 20.1 | 4.3 | 4.2 | .97 |
| MAZE: | | | | | | | |
| Treatment | 27 | 6.8 | 6.1 | 10.5 | 6.0 | 3.7 | .61 |
| Control | 17 | 7.9 | 4.2 | 11.2 | 4.7 | 3.3 | .74 |
| WJRM: | | | | | | | |
| Treatment | 30 | 24.3 | 8.2 | 26.6 | 6.4 | 2.3* | .31 |
| Control | 20 | 30.0 | 8.2 | 29.1 | 6.1 | .9 | .12 |

Note.—Cohen's *d* modeled as (posttest-pretest)/pooled *SD*.

* $p < .05$ interaction of within-groups pre- to posttest by between-groups condition.

control groups. The treatment group ($n = 27$) grew from a mean of 6.8 to 10.5 net correct words per 180 seconds. The control group ($n = 17$) grew from a mean of 7.9 to a mean of 11.2. Students with learning disabilities ($n = 5$) and English learners ($n = 2$) did not grow more on MAZE knowledge than students in the control group.

**Passage comprehension of the Woodcock Reading Mastery Tests—Revised.** The RM-ANOVA revealed a significant difference between groups, $F(1, 48) = 4.34$, $p = .04$, partial $\eta^2 = .2$, $d = .31$. The treatment group ($n = 30$) grew from a mean of 24.3 to a mean of 26.6. The control group ($n = 20$) fell from a mean of 30.0 to a mean of 29.1. Students with learning disabilities in the intervention group ($n = 5$) grew from a mean on WRMT-R/NU of 16.80 ($SD = 5.54$) to 22.40 ($SD = 2.70$). English learners ($n = 2$) in the intervention grew from a mean on WRMT-R/NU of 17.50 ($SD = 2.00$) to 24.30 ($SD = 4.00$).

**English learners.** None of the students with learning disabilities across the two studies were also considered non-English proficient, according to the CELDT testing. English learners across the two studies ($n = 5$) gained an average of 15 wpm. From the anecdotal records of the credential candidates serving as instructors, we deduced that, without exception, students who read 35 wpm or less (a first-grade level) in either the control or intervention groups were students who appeared to be weaker in their English language skills (though the designation by the school did not indicate differences). Of the five students who were designated by the CELDT as having intermediate fluency in our two studies, three were those who read 35 wpm or less. Anecdotal records indicate that the others who had this intermediate fluency designation appeared to be stronger in English than the CELDT test indicated. This is further evidence that for the older students, the English learner designation is very muddy and confusing; that is, in our work we have found that all students who are struggling readers—whether designated English learners or not—need academic language and content-area language development.

Anecdotal records indicated that instructors spent about 10 minutes per day on vocabulary instruction whether students were designated English learners or not. However, data indicated that when students read lower than a reading level of 2.0, the words were in Category 1 (see Beck et al., 2008, for full detail) or part of basic con-

Table 3. Combined Results of Study 1 and Study 2: Growth on Three Reading Measures in Common

| | | Pretest | | Posttest | | | |
|---|---|---|---|---|---|---|---|
| | *n* | Mean | *SD* | Mean | *SD* | Change | *d* |
| ORF: | | | | | | | |
| Treatment | 61 | 88.5 | 31.5 | 108.3 | 32.4 | 19.8* | .62 |
| Control | 48 | 102.5 | 30.2 | 103.4 | 32.4 | .9 | .03 |
| Vocabulary: | | | | | | | |
| Treatment | 60 | 14.2 | 3.5 | 17.5 | 4.4 | 3.3 | .83 |
| Control | 44 | 15.7 | 4.5 | 19.6 | 4.5 | 3.9 | .87 |
| MAZE: | | | | | | | |
| Treatment | 55 | 6.3 | 5.8 | 10.1 | 5.8 | 3.8 | .66 |
| Control | 41 | 8.7 | 5.7 | 11.9 | 6.6 | 3.2 | .54 |

Note.—Cohen's *d* is modeled as (posttest-pretest)/pooled *SD*.
* $p < .05$ interaction of within-groups pre- to posttest by between-groups condition.

versation in English. For example, one student who was a recent immigrant from Somalia asked, "What is a bin?" Another student in the group said, "Oh, I know, like I been to the store." Thus, a lesson on the words *been* and *bin* ensued, including writing definitions and sentences using both words.

When students read higher than a reading level of 2.0, vocabulary instruction took about the same amount of time during the lesson; however, the words of focus were higher-level words in English, including both academic language vocabulary (Beck's Category 2 words) and content-focused vocabulary (Beck's Category 3 words). For example, during one lesson the instructor said, "Let's talk about the author's point of view." One student asked, "What is that?" Instruction ensued on this academic language vocabulary, including examples, a definition, and the use of the concept in sentences. The instructor and the students then returned to the text to discuss the author's point of view.

## Combined Study 1 and Study 2

**Oral reading fluency.** The RM-ANOVA revealed that growth in mean ORF scores within the intervention group across both studies was significant and depended on receiving the intervention, $F(1, 107) = 140.56$, $p < .05$, partial $\eta^2 = .51$, $d = .62$. ORF rates grew for both the treatment and control groups (see Table 3). The treatment group ($n = 61$) grew from a mean of 88.5 to 108.3, a gain of about 20 wpm over the intervention period. In contrast, the control group ($n = 48$) grew by about 1 wpm during the same period, from a mean of 102.5 to 103.4.

**Vocabulary.** RM-ANOVA indicated no differences between groups on the vocabulary measure, $F(1, 102) = .43$, $p = .516$, partial $\eta^2 = .04$. The number of words correctly identified grew for both the treatment and control groups in both studies. The treatment group ($n = 60$) grew from a mean of 14.6 to a mean of 17.5 on the 25-item test. The control group ($n = 44$) grew from a mean of 15.7 to 19.6.

**MAZE comprehension.** RM-ANOVA indicated no differences between groups on the MAZE measure across the two studies, $F(1, 94) = 3.25$, $p = .08$, partial $\eta^2 = .03$. The number of correctly identified words in the MAZE probe grew for both the treatment and control groups. The treatment group ($n = 55$) grew from a mean of 6.3

Table 4. Study 1: Growth on Three Reading Measures by Reading Level

| | | Pretest | | Posttest | | | |
|---|---|---|---|---|---|---|---|
| | *n* | Mean | *SD* | Mean | *SD* | Change | *d* |
| ORF: | | | | | | | |
| Low: | | | | | | | |
| Treatment | 10 | 49.4 | 9.7 | 70.5 | 15.2 | 21.1 | 1.7 |
| Control | 3 | 29.0 | 13.9 | 29.0 | 14.8 | 0 | 0 |
| Medium: | | | | | | | |
| Treatment | 15 | 98.5 | 10.4 | 115.1 | 13.2 | 16.6 | 1.4 |
| Control | 16 | 98.6 | 11.6 | 100.4 | 10.5 | 1.8 | .2 |
| High: | | | | | | | |
| Treatment | 6 | 129.6 | 13.3 | 146.3 | 17.8 | 16.7 | 1.1 |
| Control | 9 | 132.7 | 11.1 | 133.1 | 13.9 | .4 | 0 |
| Vocabulary: | | | | | | | |
| Low: | | | | | | | |
| Treatment | 10 | 11.9 | 2.8 | 11.0 | 1.8 | .9 | .4 |
| Control | 3 | 7.7 | 4.5 | 12.6 | 2.0 | 4.9 | 1.4 |
| Medium: | | | | | | | |
| Treatment | 14 | 15.4 | 3.7 | 18.8 | 4.0 | 3.4 | .9 |
| Control | 15 | 15.1 | 3.1 | 18.9 | 4.0 | 3.8 | 1.1 |
| High: | | | | | | | |
| Treatment | 6 | 14.3 | 3.1 | 21.3 | 1.5 | 7.0 | 2.9 |
| Control | 9 | 19.1 | 2.5 | 22.7 | 1.2 | 3.6 | 1.8 |
| MAZE: | | | | | | | |
| Low: | | | | | | | |
| Treatment | 9 | 0.8 | 5.7 | 4.6 | 3.4 | 3.8 | .8 |
| Control | 1 | 4.0 | ... | 8.0 | ... | 4.0 | ... |
| Medium: | | | | | | | |
| Treatment | 14 | 7.6 | 3.0 | 12.2 | 5.0 | 4.6 | 1.1 |
| Control | 13 | 7.5 | 6.4 | 11.0 | 7.4 | 3.5 | .5 |
| High: | | | | | | | |
| Treatment | 6 | 9.1 | 5.2 | 12.0 | 5.3 | 2.9 | .6 |
| Control | 9 | 13.3 | 3.4 | 16.7 | 3.0 | 3.4 | 1.1 |

to 10.1 net correct words per 180 seconds. The control group ($n = 41$) grew from a mean of 8.7 to 11.9.

## Post Hoc Research Question

After the study, we asked the question, do students who are relatively high, medium, and low performers within the sample of students who are far below and below basic in sixth grade demonstrate different patterns of performance? It seemed important to examine the responses of low, middle, and high performers within both of our studies. Conceptually, it made sense due to grade-level norms to divide the groups as follows: a low group, reading at 70 wpm or lower (an instructional reading level of 2.0 or lower); a middle group, reading at 71 to 115 wpm (an instructional reading level of about 2.5 to 3.0); and a high group, reading at 116 to 150 wpm (an instructional reading level of about 3.0 to 3.5). However, we did not initially stratify by group level.

**Study 1.** Results indicate very strong effect sizes on ORF (see Table 4) for low, middle, and high group intervention students ($d = 1.7$, 1.4, and 1.1, respectively). Attrition in the lowest group of the control condition rendered the effect size for this

Table 5. Study 2: Growth on Four Reading Measures by Reading Level

| | | Pretest | | Posttest | | | |
|---|---|---|---|---|---|---|---|
| | *n* | Mean | *SD* | Mean | *SD* | Change | *d* |
| ORF: | | | | | | | |
| Low: | | | | | | | |
| Treatment | 10 | 53.4 | 15.1 | 75.3 | 25.6 | 21.9 | 1.0 |
| Control | 2 | 43.0 | 7.1 | 43.5 | 6.4 | .5 | .1 |
| Medium: | | | | | | | |
| Treatment | 13 | 93.7 | 13.6 | 115.3 | 17.1 | 21.6 | 1.4 |
| Control | 12 | 99.7 | 13.6 | 101.1 | 14.3 | 1.4 | .1 |
| High: | | | | | | | |
| Treatment | 7 | 128.0 | 9.3 | 149.1 | 11.0 | 21.1 | 2.1 |
| Control | 6 | 130.2 | 13.5 | 128.9 | 18.2 | 1.3 | .1 |
| Vocabulary: | | | | | | | |
| Low: | | | | | | | |
| Treatment | 9 | 11.4 | 2.9 | 14.2 | 4.4 | 2.8 | .8 |
| Control | 2 | 10.0 | 7.1 | 11.0 | 1.4 | 1.0 | .2 |
| Medium: | | | | | | | |
| Treatment | 13 | 14.8 | 3.1 | 18.6 | 3.4 | 3.8 | 1.2 |
| Control | 12 | 15.7 | 3.5 | 20.2 | 2.7 | 4.5 | 1.4 |
| High: | | | | | | | |
| Treatment | 7 | 16.6 | 2.8 | 20.7 | 1.8 | 4.1 | 1.7 |
| Control | 4 | 19.5 | 3.3 | 24.0 | .8 | 4.5 | 1.9 |
| MAZE: | | | | | | | |
| Low: | | | | | | | |
| Treatment | 8 | 1.6 | 6.3 | 5.0 | 5.8 | 3.4 | 4.1 |
| Control | 2 | 4.5 | .7 | 3.0 | 1.4 | 1.5 | 1.4 |
| Medium: | | | | | | | |
| Treatment | 13 | 7.9 | 3.8 | 11.9 | 4.8 | 4 | .9 |
| Control | 12 | 7.4 | 3.8 | 11.9 | 4.0 | 4.5 | 1.2 |
| High: | | | | | | | |
| Treatment | 6 | 11.3 | 4.0 | 15.3 | 2.7 | 4 | 1.2 |
| Control | 3 | 12.0 | 4.6 | 14.0 | 1.7 | 2 | .6 |
| WRMT: | | | | | | | |
| Low: | | | | | | | |
| Treatment | 10 | 20.5 | 6.9 | 21.9 | 4.9 | 1.4 | .2 |
| Control | 2 | 16.0 | 5.7 | 19.0 | 12.8 | 3 | .3 |
| Medium: | | | | | | | |
| Treatment | 13 | 24.4 | 7.2 | 26.9 | 6.1 | 2.5 | .4 |
| Control | 12 | 29.0 | 5.7 | 29.1 | 4.5 | .1 | 0 |
| High: | | | | | | | |
| Treatment | 7 | 29.4 | 5.9 | 32.9 | 5.4 | 3.5 | .6 |
| Control | 6 | 36.5 | 7.0 | 32.5 | 3.4 | 4.0 | .7 |

group questionable. The middle group was least affected by attrition and appeared to be relatively equal in size in both the control and the treatment groups. The effect sizes for the middle group in ORF and MAZE comprehension were strong ($d = 1.4$ and 1.1, respectively).

**Study 2.** As in Study 1, results for the low, middle, and high intervention students (see Table 5) indicate very strong effect sizes for ORF ($d = 1.0$, 1.4, and 2.1, respectively). The MAZE low-group effect size was very strong ($d = 4.1$); however, once again, attrition in the low group of the control condition makes this effect size questionable. The middle group WRMT-R/NU effect size was moderate ($d = .40$).

## Discussion

Studies 1 and 2 were designed to investigate intensive Tier II intervention in literacy development for those at a third-grade reading level or lower in the sixth grade. Our most important and substantive findings are significant differences between treatment and control on speed and accuracy while reading text ($d = .62$) and on reading comprehension ($d = .40$). Those who received the intervention benefited, including students with learning disabilities and English learners. Post hoc separation of low, middle, and high readers into groups (within the low-performing sixth-grade sample) yielded very strong effect sizes when viewed separately. The lowest readers received decoding instruction in Corrective Reading A or B1. Students in the middle and high groups all received instruction in REWARDS. Students in low, middle, and high groups all received instruction in leveled reading via Read Naturally and reading comprehension/vocabulary instruction as part of the Daybook.

In Study 1, students in the treatment condition gained about 18 wpm in 10 weeks ($d = .57$); students with learning disabilities gained about 22 wpm. In Study 2, students who received the intervention gained 21.6 wpm ($d = .66$). Students with learning disabilities in the intervention group gained 14 wpm and gained 6 points on the WRMT-R/NU; these students also demonstrated gains, though not significant, on the MAZE measure. Students with learning disabilities gained an average of 18 wpm across the two studies ($n = 10$), and English learners across the two studies ($n = 5$) gained an average of 15 wpm. These wpm gains exceeded the one-word gain per week expectation for oral reading progress (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Silberglitt & Hintze, 2007).

In post hoc analyses, when we restricted the variance by creating groups of low, medium, and high readers within the struggling-reader group, we appeared to obtain a more accurate picture of the strength of the intervention. Our effect sizes appear to be very strong and can be contrasted with previous studies showing very small impacts on fluency, particularly in older struggling readers (Spencer, 2010; Torgesen et al., 2001; Vaughn et al., 2010). The MAZE and vocabulary results were discouraging, however, and perhaps our results may indicate that more sensitive and revealing measures in these areas are necessary. Instructor records yielded details about vocabulary instruction and other rich data that can inform future instructional designing and planning for Tier II instruction in sixth grade.

Previous studies of Read Naturally have reported no significant differences on alphabetics, fluency, comprehension, and general reading for participants in the first and second grades (Denton et al., 2008), with effect sizes not large enough to be judged substantively meaningful. In our studies, we investigated whether Read Naturally materials could be used effectively to improve fluency in sixth-grade struggling readers when used as one component in an intensive intervention bundle. We selected it because it provided a set of leveled reading materials allowing for abundant oral reading practice at the students' precise reading level (O'Connor, Swanson, & Geraghty, 2010; Vadasy, Sanders, & Peyton, 2005). While repeated reading was a part of the curriculum, students read a new passage daily. Oral reading for speed and accuracy was possible with leveled materials readily available for use. Fluency-building work appears to hold promise as one of the critical pieces for reading comprehension development (Yovanoff, Duesbery, Alonzo, & Tindal, 2005).

## Limitations

In both studies, attrition presented a threat to causal validity, since higher achievers show faster rates of change compared to lower-achieving students. By the end of the 10 weeks in Study 1, 31 of the 33 students remained in the treatment condition (five had learning disabilities); 28 remained in the control condition. In Study 2, all 30 students remained in the treatment condition (five had learning disabilities); only 20 remained in the control condition (four had learning disabilities). There was even more attrition in the vocabulary and MAZE measures, as during the testing period many students were absent or had moved and we were not able to locate them. In addition, our measurement instruments were limited, particularly in assessing comprehension and vocabulary growth for middle school readers when reading levels range from early first to middle third grade. As discussed previously, this created a range of variance within the group of struggling readers that not only led to large standard deviations but also to potential measurement flaws.

Our studies had limited sample sizes, and both took place in only one school, which may have, for some reason, been ideal for positive effects of Tier II intervention (Mellard, Byrd, Johnson, Tollefson, & Boesche, 2004). While we noted effects with the use of our bundle of components at this school site, we do not know which components contributed the most to the progress students made. That is, we are uncertain of which of our instructional materials contributed the most to the ORF and WRMT-R/NU gains. Hence, our work warrants further investigation—more specifically, a component analysis study would potentially illuminate the field.

## Educational Implications and Future Research

Our results seem to indicate that Tier II instruction with an equal balance on word analysis, fluency building, comprehension, and vocabulary instruction, combined with evidence-based Tier I, has a significant impact on struggling readers in sixth grade. Our work suggests that the Tier II instructional supplement should be an option for struggling readers in secondary schools with the goal of effective literacy development. Continually stronger and perhaps longer Tier II interventions are possible and need to be designed, perhaps across grade levels, particularly in light of the findings of secondary meta-analyses in reading comprehension (Edmonds et al., 2009).

Older struggling readers deserve the best-designed interventions. If intensive instruction can be designed to produce significant changes in the reading abilities of students in sixth grade, it may potentially lead to the prevention of early school exit, excessive referrals for learning disabilities, and low self-esteem due to lack of success. Clarity regarding who could benefit from Tier II instruction and how Tier III instructional services should be implemented are matters to be grappled with by general and special educators in the twenty-first century (Scruggs & Mastropieri, 2002). Studies like those described here are important in the quest to define and develop Tier II interventions for students who are older struggling readers. The development of successful RTI models or other models that make strides toward closing the achievement gap can become the lifeline that all students deserve.

## Note

This research was partially supported by grant H325K060356 from the U.S. Department of Education, Office of Special Education Projects (OSEP).

## References

Alonzo, J., & Tindal, G. (2009). *Alternate form and test-retest reliability of EasyCBM reading measures* (Technical Report no. 0906). Eugene: Behavioral Research and Teaching, University of Oregon.

Archer, A., Gleason, M., & Vachon, V. (2002). *REWARDS (Reading excellence: Word attach & rate development strategies).* Longmont, CO: Sopris West.

Archer, A. L., Gleason, M. M., & Vachon, V. L. (2003). Decoding and fluency: Foundation skills for struggling older readers. *Learning Disability Quarterly,* **26**(10), 89–101.

Beck, I. L., McKeown, M. G., & Kucan, L. (2008). *Creating robust vocabulary: Frequently asked questions and extended examples.* New York: Guilford.

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S., Shaywitz, B., & Shaywitz, S. (2004). Effects of intensive reading remediation for second and third graders. *Journal of Educational Psychology,* **96**, 444–461.

Bradley, R., & Danielson, L. (2004). The Office of Special Education Program's LD initiative: A context for inquiry and consensus. *Learning Disability Quarterly,* **27**(4), 186–188.

Coyne, M. D., Kame'enui, E. J., & Simmons, D. C. (2004). Beginning reading intervention as inoculation or insulin: First-grade reading of strong responders to kindergarten intervention. *Journal of Learning Disabilities,* **37**(2), 90–104.

Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education,* **37**(3), 184–192.

Denton, C. A., Fletcher, J. M., Anthony, J. L., & Francis, D. J. (2006). An evaluation of intensive intervention for students with persistent reading difficulties. *Journal of Learning Disabilities,* **39**(5), 447–466.

Denton, C. A., Wexler, J., Vaughn, S., & Bryan, D. (2008). Intervention provided to linguistically diverse middle school students with severe reading difficulties. *Learning Disabilities Research and Practice,* **23**(2), 79–89.

EasyCBM. (2009). Eugene, OR: Behavioral Research and Teaching.

Edmonds, M. S., Vaughn, S., Wexler, J., Reutenbuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects of reading comprehension outcomes on older struggling readers. *Review of Educational Research,* **79**(1), 262–300.

Engelmann, S., Meyer, L., Carnine, L., Becker, W., Eisele, J., & Johnson, G. (1999). *Corrective reading, decoding.* Columbus, OH: McGraw-Hill.

Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia,* **54**(2), 304–331.

Fletcher, J. M., Foorman, B. R., Boudousquie, A., Barnes, M. A., Schatschneider, C., & Francis, D. J. (2002). Assessment of reading and learning disabilities: A research-oriented approach. *Journal of School Psychology,* **40**(1), 27–63.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly,* **41**(1), 93–99.

Fuchs, L. S., Fuchs, D., Hamlett, C. S., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review,* **22**(1), 27–48.

Gersten, R., Baker, S., Haager, D., & Graves, A. W. (2005). Exploring the role of teacher quality in predicting reading outcomes for first grade English learners: An observational study. *Remedial & Special Education,* **38**, 212–222.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades* (NCEE 2009-4045). Washington, DC: National

Center for Education Evaluation Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Graves, A. W. (2010). The longitudinal study of the impact of effective beginning reading instruction for English learners: Literacy, language, and learning disabilities. In T. Scruggs & M. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (vol. **23**, pp. 155–174). London: Emerald.

Graves, A. W., Brandon, R., Duesbery, L., McIntosh, A., & Block, N. F. (2007). *The Tier II intervention in sixth grade: A pilot study* (Technical report no. 39). Project Achieve, San Diego State University.

Graves, A. W., Gersten, R., & Haager, D. (2004). Literacy instruction in multiple-language first-grade classrooms: Linking student outcomes to observed instructional practice. *Learning Disabilities Research & Practice,* **19**(4), 262–272.

Graves, A. W., Plasencia-Peinado, J., Deno, S., & Johnson, J. (2005). Formatively evaluating the reading progress of first-grade English learners in multiple language classrooms. *Remedial & Special Education,* **38**, 245–255.

Greenwood, C. R., Tapia, Y., Abbott, M., & Walton, C. (2003). A building-based case study of evidence-based literacy practices: Implementation, reading behavior, and growth in reading fluency, K–4. *Journal of Special Education,* **37**, 200–215.

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *Journal of Special Education,* **34**, 90–103.

Ihnot, C., Matsoff, J., Gavin, J., & Hendrickson, L. (2001). *Read naturally.* St. Paul, MN: Read Naturally.

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446 (2006).

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *Reading Teacher,* **59**(7), 635–644.

Lloyd, J., Cullinan, D., Heins, E., & Epstein, M. (1980). Direct instruction: Effects on oral and written language comprehension. *Learning Disability Quarterly,* **3**(1), 70–76.

McIntosh, A., Graves, A., & Gersten, R. (2007). The effects of response to intervention on literacy development in multiple language settings. *Learning Disability Quarterly,* **30**(3), 197–212.

Mellard, D. F., Byrd, S. E., Johnson, E., Tollefson, J. M., & Boesche, L. (2004). Foundations and research on identifying model responsiveness-to-intervention sites. *Learning Disability Quarterly,* **27**(4), 243–256.

National Governor's Association. (2009). *Implementing graduation counts.* Washington, DC: National Governor's Association Center for Best Practices.

O'Connor, R. E., Harty, K. R., & Fulmer, D. (2005). Tiers of intervention in kindergarten through third grade. *Journal of Learning Disabilities,* **38**, 532–538.

O'Connor, R. E., Swanson, L. H., & Geraghty, C. (2010). Improvement in reading rate under independent and difficult text levels: Influences on word and comprehension skills. *Journal of Educational Psychology,* **102**(1), 1–19.

Purcell-Gates, V., Duke, N. K., & Martineau, J. A. (2007). Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching. *Reading Research Quarterly,* **42**(1), 8–45.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutenbuch, C. K., & Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice.* Portsmouth, NH: Research Corp.

Schleppegrell, M. J., & Go, A. (2007). Analyzing the writing of English learners: A functional approach. *Language Arts,* **84**(6), 529–538.

Scruggs, T. E., & Mastropieri, M. A. (2002). On babies and bathwater: Addressing the problems of identification of learning disabilities. *Learning Disabilities Quarterly,* **25**(3), 155–164.

Shinn, M., Deno, S., & Espin, C. (2000). Technical adequacy of the MAZE task for curriculum-based measurement of reading growth. *Journal of Special Education,* **34**(3), 164–172.

Silberglitt, B., & Hintze, J. M. (2007). A conditional analysis of R-CBM growth rates by level of performance. *Exceptional Children,* **74**(1), 71–84.

Spandel, V., Nathan, R., & Robb, L. (2001). *Daybook for critical reading and writing.* Wilmington, MA: Great Source Education Group.

Spencer, A. (2010). The effects of a fluency intervention program on the fluency and comprehension outcomes of middle-school students with severe reading deficits. *Learning Disabilities Research & Practice,* **25**(10), 75–86.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities,* **34**(1), 33–58.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2005). Relative effectiveness of reading practice or word-level instruction in supplemental tutoring: How text matters. *Journal of Learning Disabilities,* **38**, 364–380.

Vaughn, S., Cirino, P. T., Wanzek, J., Wexler, J., Fletcher, J. M., Denton, C. A., . . . Francis, D. J. (2010). Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review,* **39**(1), 3–21.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children,* **69**, 391–409.

Vaughn, S., Linan-Thompson, S., Kouzekannani, K., Bryant, D. P., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education,* **24**, 127–141.

Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention: Examining higher and lower responders. *Exceptional Children,* **75**, 165–183.

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing,* **21**, 437–480.

Woodcock, R. W. (2003). *Woodcock reading mastery tests—revised normative update (WRMT-R/NU).* Boston: Pearson.

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice,* **24**(3), 4–12.