

This article was downloaded by: [. Carolyn A. Denton]

On: 06 July 2011, At: 05:05

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Journal of Research on Educational Effectiveness

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/uree20>

### An Experimental Study of Scheduling and Duration of "Tier 2" First-Grade Reading Intervention

Carolyn A. Denton<sup>a</sup>, Paul T. Cirino<sup>b</sup>, Amy E. Barth<sup>b</sup>, Melissa Romain<sup>b</sup>, Sharon Vaughn<sup>c</sup>, Jade Wexler<sup>c</sup>, David J. Francis<sup>b</sup> & Jack M. Fletcher<sup>b</sup>

<sup>a</sup> University of Texas Health Science Center Houston, Houston, Texas, USA

<sup>b</sup> University of Houston, Houston, Texas, USA

<sup>c</sup> University of Texas, Austin, Texas, USA

Available online: 05 Jul 2011

To cite this article: Carolyn A. Denton, Paul T. Cirino, Amy E. Barth, Melissa Romain, Sharon Vaughn, Jade Wexler, David J. Francis & Jack M. Fletcher (2011): An Experimental Study of Scheduling and Duration of "Tier 2" First-Grade Reading Intervention, *Journal of Research on Educational Effectiveness*, 4:3, 208-230

To link to this article: <http://dx.doi.org/10.1080/19345747.2010.530127>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## An Experimental Study of Scheduling and Duration of “Tier 2” First-Grade Reading Intervention

**Carolyn A. Denton**

University of Texas Health Science Center Houston, Houston, Texas, USA

**Paul T. Cirino, Amy E. Barth, and Melissa Romain**

University of Houston, Houston, Texas, USA

**Sharon Vaughn and Jade Wexler**

University of Texas, Austin, Texas, USA

**David J. Francis and Jack M. Fletcher**

University of Houston, Houston, Texas, USA

**Abstract:** This study compared the effects on reading outcomes of delivering supplemental, small-group intervention to first-grade students at risk for reading difficulties randomly assigned to one of three different treatment schedules: extended (4 sessions per week, 16 weeks;  $n = 66$ ), concentrated (4 sessions per week, 8 weeks;  $n = 64$ ), or distributed (2 sessions per week, 16 weeks;  $n = 62$ ) schedules. All at-risk readers, identified through screening followed by 8 weeks of oral reading fluency (ORF) progress monitoring, received the same Tier 2 reading intervention in groups of 2 to 4 beginning in January of Grade 1. Group means were higher in word reading and ORF at the final time point relative to pretest; however, the groups did not differ significantly on any reading outcome or on rates of adequate intervention response. Of potential covariates, site, age, free lunch status, program coverage rate, and tutor were significantly related to student outcomes; however, the addition of these variables in multivariate models did not substantially change results. Rates of adequate intervention response were lower than have been reported for some first-grade interventions of longer duration.

**Keywords:** Reading intervention, first grade, duration, massed and distributed practice

The Response to Intervention (RTI) framework is a school-wide initiative designed to promote positive achievement and behavioral outcomes for all students (Glover, 2010). School districts across the United States are increasingly implementing RTI models (Berkeley, Bender, Peaster, & Saunders, 2009; Harr-Robins, Shambaugh, & Parrish, 2009). Although RTI may address a variety of academic and behavior concerns, existing implementations often focus on reading difficulties (Spectrum K12 School Solutions, 2009), which is the focus of this article. A salient characteristic of RTI is the implementation of a multitiered service delivery system based on the analysis of student assessment data. Commonly, RTI reading models include three tiers of intervention, in which Tier 1 consists of universal

Address correspondence to Carolyn A. Denton, University of Texas Health Science Center Houston, Pediatrics, Children’s Learning Institute, 7000 Fannin, UCT 2443, Houston, TX 77030, USA. E-mail: Carolyn.A.Denton@uth.tmc.edu

screening and progress monitoring, and quality classroom reading instruction provided to all students using research-validated materials and approaches. This instruction is differentiated to address the needs of groups of students within the classroom, and in some models includes systematic instructional adaptations (Kovaleski & Black, 2010). Students who do not make adequate progress in Tier 1 are provided supplemental small-group Tier 2 intervention; those with inadequate response in Tier 2 receive more intensive Tier 3 intervention. Although RTI reading models are widely implemented, some key questions remain. One concerns the ideal intensity of interventions at Tiers 2 and 3 (Gansle & Noell, 2007; Ikeda et al., 2007).

### SCHEDULING AND DURATION OF READING INTERVENTIONS

Tier 2 must be provided with sufficient intensity to “powerfully accelerate development” of reading skills and “prevent reading problems for most students” (Al Otaiba & Torgesen, 2007, p. 213–214). The intensity of an intervention is affected by several factors, including the amount of time devoted to intervention in the weekly schedule and the number of weeks for which it is provided. The amount of time devoted to Tier 2 intervention varies across RTI implementations. In some cases, Tier 2 is operationalized as a relatively brief intervention. For example, Marston, Lau, and Muyskens (2007) described an RTI service delivery model in which Tier 2 interventions are generally provided over an 8-week period and may consist of the use of published reading programs or more general types of academic support (e.g., small reading groups, peer tutoring). In other approaches, including many that have been evaluated by researchers, supplemental reading interventions were provided for 20 weeks or more (Wanzek & Vaughn, 2007). Based on characteristics of Tier 2 reading interventions with demonstrated efficacy, the U.S. Department of Education’s What Works Clearinghouse (Gersten et al., 2008) recommended that educators provide Tier 2 reading intervention three to five times per week for 20 to 40 min in addition to regular classroom reading instruction. The What Works Clearinghouse report concluded that inferences regarding the number of weeks for which intervention should be provided were not possible based on the reviewed studies, recommending that Tier 2 be provided “for a reasonable amount of time before providing a more intensive daily Tier 3 intervention” (Gersten et al., 2008, p. 26).

Research provides limited insight into questions related to intervention dosage and scheduling in the primary grades, and results have been mixed. For example, Wanzek and Vaughn (2008) observed in two parallel studies that providing 30 min of daily intervention in the fall of first grade followed by 60 min of daily intervention in the spring did not appear to increase the number of students with adequate instructional response relative to providing 30 min per day throughout the school year. Hatcher et al. (2006) similarly found that Year 1 British students with reading difficulties who received supplemental intervention for two consecutive 10-week periods (33 hr of instruction) performed comparably to a group who received the same intervention only during the second 10-week period (16.5 hr of instruction). In contrast, Al Otaiba, Schatschneider, and Silverman (2005) found differences favoring more extended intervention for kindergarten students randomly assigned to receive a small-group intervention for 30 min either two or four times per week or to a control condition. Students in the four-times-per-week group significantly outperformed controls in word reading and comprehension, with large effect sizes, whereas those who received intervention two times per week performed significantly better than controls on only one measure of phonemic awareness.

Student outcomes may also be affected when intervention is provided on a more concentrated schedule over a shorter period (e.g., 4 days per week for 10 weeks) or a less concentrated schedule over a longer period (e.g., 2 days per week for 20 weeks). It has long been established that spacing the presentation and practice of items over time (i.e., distributed) is more effective than presenting or practicing large amounts of content in fewer sessions (i.e., massed) for verbal learning tasks (e.g., Underwood, 1961), an effect that is especially salient for the retention of learned information (Fishman, Keller, & Atkinson, 1968; Smith & Rothkopf, 1984). However, studies of the application of this principle to scheduling reading instruction in schools are few and inconclusive. In two quasi-experimental studies, Seabrook, Brown, and Solity (2005) found that young children performed better in letter-sound correspondences and word reading after receiving three 2-min instructional sessions each day relative to students who received one 6-min daily session, whereas Ukrainetz, Ross, and Harm (2009) found few differences in outcomes related to delivering phonemic awareness intervention to at-risk kindergarten students three times per week for 3 months versus once a week for 6 months.

## STUDY PURPOSE AND RESEARCH QUESTIONS

There are significant gaps in the research base related to the optimal duration and scheduling of supplemental reading interventions. The reviewed research is limited, and findings have been mixed. The purpose of this study was to compare the effects of Tier 2 intervention provided to first-grade students at risk for reading difficulties on three different schedules, each designed to mimic the levels of intensity often provided by schools implementing RTI models.

We addressed the following research questions:

- RQ1: Do first-grade students at risk for reading difficulties differ in reading outcomes following the same Tier 2 intervention provided in 4 sessions per week for 16 weeks (extended schedule), 4 sessions per week for 8 weeks (concentrated schedule), or 2 sessions per week for 16 weeks (distributed schedule)?
- RQ2: Do students who receive intervention on these schedules differ in their rates of adequate intervention response?

We hypothesized that student outcomes and adequate instructional response rates would be higher for the extended schedule group. We also hypothesized that the distributed schedule would be associated with better outcomes than the concentrated schedule.

## METHOD

### Participants

*School Sites.* This study was conducted in nine schools located in two school districts in the southwestern United States; four were part of a large urban district, and five were in a smaller, partly rural district. The Institutional Review Boards from each of the participating universities and the research review committees of the participating school districts approved this research. All schools met minimum state accountability standards. In the larger district, populations in participating schools were primarily African American (51%) and

Hispanic (35%), with 9% White and 5% other ethnicities, and the percentage of students with economic disadvantage ranged from 62% to 97% ( $M = 80\%$ ,  $SD = 17$ ). In the smaller district, students were primarily Hispanic (79%), with 12% African American, 8% White, and 1% other ethnicities; the rates of economic disadvantage ranged from 79% to 93% ( $M = 83\%$ ,  $SD = 6$ ).

*Criteria for Participation.* We identified first-grade students as at risk for reading difficulties based on a two-step process. First, we screened all students in the 1st month of Grade 1, using a brief screen from the Texas Primary Reading Inventory (TPRI; Foorman, Fletcher, & Francis, 2004); only students who did not meet criteria on this screen were further considered for intervention. The TPRI screen is composed of items predictive of adequate first-grade reading development; students who passed the screen were able to (a) provide 8 of 10 letter-sound correspondences, (b) read at least three words from a list of words typically learned in first grade, and (c) blend three to five separately pronounced phonemes to identify words. Because the TPRI is designed to minimize false negative errors (i.e., to not miss students who might need additional help), the false positive identification rate is relatively high (i.e., some students who fail the screen do not actually encounter reading difficulties). Therefore, to ensure the at-risk status of the students in this study, we monitored the oral reading fluency (ORF) progress of students who had not passed the TPRI screen every 2 weeks for 8 weeks. Students were considered at risk and qualified for intervention if, at the fourth time point, they read fewer than 15 words correct per minute (wcpm) and had also read fewer than 10 wcpm on at least one of the first three time points. These cut-points were selected because they approximate the median scores at each of the four progress monitoring time points.

*Student Participants.* The preliminary sample consisted of 680 first-grade students who were screened across the nine schools (397 in the small district and 283 in the large district). Students were excluded from the study only if they received their primary reading instruction outside of the general education classroom or in a language other than English or if they had school-identified severe intellectual disabilities or severe emotional disturbance. Of the 680 students, 219 passed the TPRI screen and were considered typical readers, whereas 461 did not pass the screen. Of these 461 students, 273 continued to meet at-risk criteria after four waves of progress monitoring and thus qualified for intervention.

The 273 at-risk students were randomly assigned to one of three treatment conditions reflecting relatively more extended, concentrated, and distributed intervention schedules: (a) four sessions per week for 16 weeks (extended group;  $n = 91$ ), (b) four sessions per week for 8 weeks (concentrated group;  $n = 90$ ), or (c) two sessions per week for 16 weeks (distributed group;  $n = 92$ ). In the smaller site, 117 of the 172 at-risk students were randomly selected for intervention, with the remainder designated as alternates in their respective experimental conditions, leaving 73, 72, and 73 intervention students in the extended, concentrated, and distributed groups, respectively. Across groups, some students left the study because they moved out of the district prior to intervention ( $n = 14$ ) or during intervention ( $n = 8$ ), or were removed by schools because of scheduling conflicts ( $n = 13$ ) or due to parent withdrawal ( $n = 1$ ). In some cases, students in the smaller district who left prior to pretesting were replaced by alternates who had been randomized to their same condition, if sufficient alternates were available in their schools ( $n = 26$ ), although 10 of these moved or could not be scheduled. Because the purpose of this study was to evaluate the impact of various intervention dosages and schedules, we also excluded from final analyses any students who received insufficient intervention due to excess absences

( $n = 5$ ) even though they were posttested. Extended group students had to attend at least 40 sessions to be included, and concentrated and distributed students had to attend at least 20 sessions. There was also one error in randomization (student inadvertently served in a group to which the student was not assigned). After these changes, the extended group was reduced from 73 to 66, the concentrated group was reduced from 72 to 64, and the distributed group was reduced from 73 to 62. These 192 students who were in treatment, received a sufficient amount of intervention, and were posttested composed the analytic sample. The 5 children who did not complete enough intervention and the mis-assigned student composed an intent-to-treat sample (3 in the concentrated group, 2 in the extended group, and 1 in the distributed group), but results were not substantively different when reevaluated with these individuals.

Although we would have found it preferable to include a no-treatment comparison group in the study design, this option was not possible because the participating school districts were providing intervention to most students not served by the research team. In addition, to reduce the number of students in intervention who may not be truly at risk (i.e., false positives), we wanted to provide progress monitoring and Tier 1 coaching throughout the fall semester before beginning Tier 2 intervention. The schools would agree to this approach only if all at-risk students received intervention beginning in January. As the effectiveness of providing first-grade reading intervention has been established in several other studies (Fletcher, Lyon, Fuchs, & Barnes, 2007), we opted to provide intervention to all groups in order to better control its implementation because our questions focused specifically on the conditions under which such interventions were most beneficial.

Within each treatment group, the students in the final sample (e.g., the 64 from the concentrated group) did not differ from the remainder of their respective groups (e.g., the other 26 originally assigned to the concentrated group) who had pretest data but did not receive intervention, on most decoding and fluency variables, with one exception (a measure of ORF in passages for the distributed group,  $p < .05$ ). Across all groups, those in the final sample also did not differ from the attrition sample on decoding or fluency variables, except for Woodcock–Johnson III (WJ III) Word Attack ( $p < .02$ ). In both cases, those in the final sample performed better. In the final sample, the groups did not differ on pretest variables (all  $p > .05$ ), reflecting the randomized assignment.

Demographic characteristics by group can be found in Table 1. Each school contributed between 10 and 44 students to the total sample of 192; from 1 to 15 students came from each of 32 classrooms in the nine schools. The proportions of students in the three treatments did not differ in terms of site, gender, free or reduced lunch status, ethnicity, age, or other demographic characteristics (all  $p > .05$ ).

*Intervention Tutors.* Tier 2 intervention was provided by 14 tutors who were not certified teachers. We decided to utilize uncertified tutors in order to implement a model that would be feasible in many schools. Moreover, there is research evidence supporting the provision of Tier 2 reading intervention by uncertified paraprofessionals (Elbaum, Vaughn, Hughes, & Moody, 2000; Grek, Mathes, & Torgesen, 2003; Vadasy, Sanders, & Peyton, 2006; Vadasy, Sanders, & Tudor, 2007). One of the tutors held a master's degree, 10 had bachelor degrees, and 3 had high school diplomas with college coursework. Eleven had prior experience tutoring students in some capacity ( $M = 2.57$  years,  $SD = 3.37$ ; range = 5 months–13 years). For 3 tutors, this included prior experience tutoring students with reading difficulties. Two tutors were male; 4 were African or African American and 10 were White. All were native English speakers.

**Table 1.** Demographic data by treatment condition and total sample

Variable	Treatment Group			Total <sup>d</sup>
	Concentrated <sup>a</sup>	Extended <sup>b</sup>	Distributed <sup>c</sup>	
Age in years ( <i>M, SD</i> )	6.57 (0.38)	6.62 (0.40)	6.60 (0.44)	6.60 (0.41)
Female	50%	47%	48%	48%
Subsidized lunch <sup>e</sup>	69%	68%	66%	68%
Special education <sup>f</sup>	25%	26%	27%	26%
English second language <sup>g</sup>	16%	9%	19%	15%
African American	34%	38%	37%	36%
Caucasian	19%	18%	11%	16%
Hispanic	45%	42%	48%	46%
Other	2%	2%	3%	2%
Larger site	42%	44%	42%	43%

*Note.* Percentages are computed relative to the number of individuals in that group (e.g., 49% of the 65 students in concentrated were female). Percentages for ethnicity within a column total of 100%. None of the values reported differed by treatment group. Special education includes speech and language classification. Age = Age at the beginning of G1. Concentrated = four times per week for 8 weeks; Extended = four times per week for 16 weeks; Distributed = two times per week for 16 weeks.

<sup>a</sup>*n* = 64. <sup>b</sup>*n* = 66. <sup>c</sup>*n* = 62. <sup>d</sup>*N* = 192. <sup>e</sup>Data unavailable for 3 participants in the total sample. <sup>f</sup>Data unavailable for 12 participants in the total sample. <sup>g</sup>Data unavailable for 30 participants in the total sample.

Prior to the onset of intervention, all tutors received a 2-day training (approximately 12 hr). Tutors also attended weekly meetings through which ongoing professional development was provided, and coaching was provided to all tutors throughout the intervention period. At the larger site, 15 weekly meetings totaled 22.5 hr at year's end, and tutors received from 6 to 14 coaching sessions each. At the smaller site, 18 weekly meetings totaled approximately 24 contact hr, and tutors received from 5 to 12 coaching sessions each.

### Description of Interventions

The 192 at-risk readers in the study all received their regular classroom reading instruction in addition to Tier 2 intervention provided by the research team. Students in all three conditions received the same Tier 2 intervention, implemented using the same procedures, but delivered on three different schedules. Eleven of the 14 tutors provided the intervention to small groups of students assigned to all three experimental groups. Because of scheduling concerns and because tutors moved out of the area after the first 8 weeks of intervention, one tutor taught in only one condition, and two only taught in two conditions.

*Tier 1.* All study participants received Tier 1 classroom reading instruction throughout first grade. In all schools, students received daily classroom reading instruction using research-based programs, and screening and progress monitoring data were collected at the beginning, middle, and end of the school year. Additional progress monitoring data were provided to the schools by the research team. Teachers in the smaller district implemented a

core reading program with a strong emphasis on explicit phonics and word study instruction, supplementing it to various degrees with a second program targeting comprehension. Two schools in this district also implemented a guided reading approach in small groups (e.g., Fountas & Pinnell, 1996), whereas teachers in the other schools provided somewhat less small-group instruction. Schools in the smaller district administered fluency-based measures from the Dynamic Indicators of Basic Early Literacy Skills (Good & Kaminski, 2002) for screening and progress monitoring. In the larger district, three of the four schools implemented a core reading program with a strong emphasis on explicit phonics and word study instruction. The other school implemented a program that addressed phonics and word study, but to a lesser degree, with an increased emphasis on comprehension. Schools in this district administered the TPRI to screen for reading difficulties, monitor progress, and as a diagnostic assessment to guide instruction.

The research team supported Tier 1 reading instruction in two ways. First, we conducted repeated ORF assessment throughout the year for all participating students (every 2 weeks in the fall, monthly in the spring), providing the data to teachers and administrators. Second, we held monthly “data meetings” with all first-grade classroom teachers. At these meetings, teachers were provided with easily interpretable line graphs of their students’ ORF scores with regression lines illustrating the slope of each student’s current scores; teachers were taught to mark the graphs with year-end ORF goals and draw “aim lines” illustrating the trajectory that would be required for students to meet these goals. Teachers compared each student’s current slope and aim line to identify students who were and were not “on track” to meet goals. We then provided brief professional development sessions illustrating how teachers could adapt their classroom reading instruction for the students who were not on a trajectory to meet the goals. Teachers were also provided on-site coaching if they agreed to participate in it. In the larger district, coaches made 57 contacts with 12 classroom reading teachers over the school year, 37 of which were on-site visits and 20 of which were e-mail or telephone interactions. In the smaller district, the coach made 88 contacts with 20 teachers, 12 of which were on-site visits.

*Tier 2 Intervention.* All participants received the same Tier 2 intervention using a modification of the 1998 version of the *Read Well* program (Sprick, Howard, & Fidanque, 1998). *Read Well* was selected because it provides systematic, explicit instruction in both decoding and fluency with application in decodable text and because it has demonstrated efficacy for supporting word reading outcomes for at-risk students when delivered by uncertified preservice teachers in a relatively brief implementation (Denton, Anthony, Parker, & Hasbrouck, 2004). We modified the published *Read Well* program by adding instruction in vocabulary and reading comprehension and by creating partially scripted lessons plans to support the tutors. A sample lesson is provided in the appendix.

This intervention approach was designed to target the needs of the students, all of whom needed instruction in phonemic decoding, word recognition, and fluency. We believed it was important to incorporate vocabulary and comprehension instruction to ensure that these early readers would understand that reading is a process of making meaning from text (rather than an exercise in correctly pronouncing words). Some students in this study performed at average levels in word reading at pretest, as they were identified on the basis of fluency criteria. Fluency in beginning readers is dependent on ease of decoding increasingly complex words and the ability to recognize these words instantly at sight (Torgesen & Hudson, 2006). Children with higher word reading but impaired fluency had begun to access simple words but were not yet able to read increasingly complex words with ease and automaticity. In our implementation, initial placement into *Read Well* was based

on assessments provided with the program, so children with more advanced decoding skills started the intervention on higher *Read Well* units. To better target the needs of individual students, tutors also used ongoing assessment to determine the appropriate pacing through the program, as described below.

During each lesson, tutors provided 10 to 12 min of word-level instruction following the *Read Well* program. This included direct instruction and practice in phonemic awareness, letter-sound correspondences, blending sounds to read decodable words, fluent word reading, and high-frequency word recognition. Each *Read Well* unit focuses on a different letter-sound correspondence and introduces new irregular words, and the program includes sufficient materials for four lessons per unit. *Read Well* also includes unit tests designed to monitor student mastery in decoding and fluency. In this study, tutors administered these as both pretests and posttests for each unit; they could administer the posttest after 2 to 4 days on a unit. If the majority of students in a group (i.e., two of three, three of four) were able to demonstrate mastery after 2 or 3 days, tutors could move to the next unit, and they could skip a unit completely if a majority demonstrated mastery at pretest. Tutors integrated continued instruction and practice on specific elements (e.g., letter-sound correspondences) in subsequent lessons as needed by any student in the group, particularly those who had not reached full mastery criteria on unit tests.

Following the daily decoding instruction, tutors spent about 20 min on text reading practice, vocabulary instruction, and comprehension instruction. Students read narrative and expository text provided with the *Read Well* program. Each unit includes “solo” and “duet” stories; solo stories are read by the students alone, whereas the teacher and students alternate reading in the duet stories. The solo stories and the student-read portions of the duet stories are decodable using elements previously taught in the program, whereas the portions read by the teacher alone contain more sophisticated vocabulary and concepts than are typically found in decodable text. As students progress through the program, the proportion of teacher-read text decreases. In this study, students read solo stories repeatedly to meet fluency goals.

Before reading, tutors provided explicit instruction in two to four vocabulary words preselected by the researchers from the day’s text. Following scripted researcher-developed teaching protocols based on Beck, McKeown, and Kucan (2002), tutors (a) pronounced the word and asked students to repeat it; (b) provided a simple “student-friendly explanation” of the word (Beck et al., 2002, p. 35); (c) used the word in a context familiar to the children; (d) illustrated it through demonstrations, pictures, and/or providing examples and nonexamples; and (e) had students use the word in a familiar context. Most words were reviewed in subsequent lessons. Each lesson plan also included two or three vocabulary words that were not taught directly; teachers explained these words using simple language during the course of the reading. To support comprehension, tutors and students engaged in discussion of the text before, during, and after reading. After reading, tutors spent about 5 to 8 min on comprehension instruction using a researcher-developed protocol. For narrative text, the primary focus was story structure, while in expository text it was identifying main ideas and details.

One to 2 days per week, times for each lesson component were shortened to allow for administration of the mastery tests previously described. While tutors assessed one student, the others in the group participated in partner reading or independent repeated reading to build fluency or practiced writing high-frequency words. Finally, each day, tutors asked the students to practice reading with their parents using take-home versions of stories that had been read in previous lessons or lists of previously taught letters and words.

The intervention was provided in 30-min sessions according to the randomized schedules, in groups of two to four students with one tutor. Only 5% of the students were served in groups of two, and the treatment groups did not differ in terms of whether they were predominantly in group sizes of three or four ( $p > .05$ ). Prior to intervention, students were assessed following the *Read Well* program's procedures for determining initial placement. Although we attempted to group students homogeneously according to their program placement, this was often not possible given the fact that there were three randomized groups at every school. Normally, schools had only enough students to form one group in each condition.

*Fidelity of Implementation.* We conceptualized fidelity of implementation as the delivery of the instructional program as it was designed, including both program adherence and quality of delivery, and the consistency with which it is delivered, both within and across tutors (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). Direct observation data were collected to determine the level of fidelity of implementation for each tutor on three different instructional days, except that tutors who delivered intervention only in the 8-week condition were observed twice. Observers were trained on the fidelity protocol through co-observation of videotaped lessons, and interobserver reliability was established prior to each wave of data collection by co-observing and independently scoring live lessons; absolute agreement between observers averaged 90% across the three time points. Data were collected on adherence to the specific components of the program as well as overall quality of implementation. Program adherence was coded by rating each of the instructional components (e.g., decoding, text reading) on a 3-point Likert-type scale (1 = low, 3 = high). The mean program adherence score across intervention components and across tutors was 2.48 ( $SD = 0.26$ , range = 2.08–2.92). Quality of implementation (e.g., pacing, appropriate use of feedback) was also rated on a 3-point scale for each instructional component (1 = low, 3 = high). The mean quality score across components and across teachers was 2.19 ( $SD = 0.29$ , range = 1.73–2.67). The mean total fidelity rating (program adherence + quality) was 2.42 ( $SD = 0.25$ , range = 2.01–2.87).

*Time in Intervention.* Although the research design specified the number of hours of intervention to be delivered in each group, due to typical school circumstances (e.g., field trips), students actually received slightly less intervention than prescribed. On average, the extended group received intervention on 59.2 days ( $SD = 4.0$ ; about 29.5 hr) rather than the 64 days (32 hr) designated in the research design. The concentrated group received intervention on an average of 28.4 days ( $SD = 3.0$ ; about 14 hr), and the distributed group on an average of 29.8 days ( $SD = 2.6$ ; about 15 hr), rather than the 32 days (16 hr) designated in the design. By design, the extended group received significantly more intervention than the concentrated or distributed groups ( $p < .0001$ ). The distributed group also received significantly more intervention than the concentrated group ( $p < .05$ ), though the difference was practically small.

*Program Coverage.* It would be expected that students in the extended condition would cover more *Read Well* units than students in the other groups, as they met nearly twice as often. Indeed, students in the extended group completed significantly more units ( $M = 27.2$  *Read Well* units;  $SD = 4.24$ ) than those in the concentrated ( $M = 17.6$ ,  $SD = 3.08$ ) or distributed ( $M = 18.2$ ,  $SD = 4.21$ ) groups,  $F(2, 189) = 124.05$ ,  $p < .0001$ ; the latter groups did not differ from one another. Because the number of lessons was not twice as great in the extended condition, we also computed coverage as a ratio of lessons covered

per instructional day. Again there were group differences,  $F(2, 189) = 57.06, p < .0001$ , and again students in distributed and concentrated conditions did not differ, but students in the extended condition covered an average of .46 *Read Well* units per session ( $SD = .06$ ; range = .34–.63), which was significantly *lower* ( $p < .0001$ ) than the number of lessons covered per session for either the concentrated ( $M = .61, SD = .08$ ; range = .49–.77) or distributed conditions ( $M = .61, SD = .14$ ; range = .39–.85).

### Additional Instruction

Classroom teachers were interviewed in December and May to document the amount of school-provided reading instruction participating students received in addition to regular classroom reading instruction. Of the 192 students who received the research intervention, 91 also received an average of 20.7 hr ( $SD = 19.5$ ) of additional school-provided instruction. Neither the proportion of students who received this instruction nor the number of hours received differed according to treatment group, site, or their combination (all  $p > .05$ ). Of the 91 students, 52 received additional school-provided instruction in the fall only (prior to the onset of the research intervention), 12 in the spring only (concurrent with the research intervention), and 27 in both fall and spring. The proportion of students receiving this instruction in each semester did not differ among the three treatment groups ( $p > .05$ ). The large majority of the additional instruction was delivered by classroom teachers or paraprofessionals and consisted of tutoring designed to support current instructional objectives in the regular classroom reading programs; 10 students received school-provided intervention from a reading specialist.

### Measures

We assessed outcomes in decoding, fluency, and reading comprehension. Detailed descriptions and more reliability and validity data for these measures can be found at <http://www.texasldcenter.org/outcomes/>. We assessed all participants at pretest in December, after 8 weeks of intervention (i.e., at the end of intervention for the concentrated group), and at 16 weeks. We also administered ORF passages every 2 weeks from September through January and monthly from February through May to monitor progress.

*Screening.* Students were screened using the TPRI (Foorman et al., 2004). The TPRI first-grade screen requires 3 to 5 min to administer and measures letter-sound knowledge, phoneme blending, and word reading as three separate subtests. Coefficient alphas for the three screening subtests are .88, .87, and .78, respectively.

*Decoding and Spelling.* We assessed reading accuracy for real words and pseudowords with the Letter-Word Identification and Word Attack subtests of the WJ III Tests of Achievement (Woodcock, McGrew, & Mather, 2001). For students in this study, coefficient alpha values were .89 and .84, respectively. The WJ III Spelling subtest involves orally dictated words written by the examinee, adapted for this study for group administration. The coefficient alpha for the present sample was .77.

*Fluency.* We administered the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) to assess fluency in lists of words and pseudowords.

Internal consistency for different forms of this well-standardized test exceeds .90. The TOWRE composite standard score was the dependent measure. We administered the Continuous Monitoring of Early Reading Skills (CMERS; Mathes & Torgesen, 2008) Oral Reading Fluency subtest (paper-and-pencil version) to measure passage reading rate and accuracy. Students read two passages orally, and the raw score is the total number of words read correctly in 60 s averaged over the two. All texts are written at approximately a Grade 1.7 readability level. Test-retest reliability was assessed for all first-grade students in participating schools, with high correlations over the first two screening periods ( $r = .93$ ).

*Comprehension.* WJ III Passage Comprehension is a cloze-based assessment in which students read brief sentences or passages and supply missing words. Coefficient alpha in the present sample was .81. The standard score was the primary dependent measure. We also administered the Passage Comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001), in which examinees read a passage and respond to multiple-choice questions. Coefficient alpha for GRADE comprehension for the present sample was .62, which was lower than desired. In the norming sample, the reliability of the measure was .87 to .90 for students aged 6 to 7 (Williams, 2001); the coefficient was likely lower in the current study because the sample was more impaired.

### **Analyses Plan and Preliminary Data Analyses**

Data preparation first involved the evaluation of distributional data both statistically and graphically for skewness, kurtosis, and normality, with few difficulties noted in this regard. The variables assessed at pretest were CMERS ORF and WJ III Letter Word Identification and Word Attack. For some measures, only raw scores were available, in which case these were analyzed. Otherwise standard scores were utilized; for outcomes for which multiple metrics are available (e.g., *W* scores in Woodcock tests), results were substantively similar regardless of the metric utilized, and standard scores were selected for ease of interpretation and comparison to other studies and because standard score benchmarks were applied to evaluate RTI.

There were two primary kinds of analyses used to address Research Question 1. For measures assessed three times (pretest, 8-week posttest, 16-week posttest), we employed a repeated measures approach, with time as the within-subjects factor and group (with covariates and additional factors where appropriate) as between-subjects factors. For measures assessed only at the two posttest time points, the primary models were an analysis of covariance, with separate posttest comparisons considering the most closely related pretest measures as covariates (e.g., single word decoding for WJ III Comprehension). Most analyses compared the three treatment groups with one another. Then each model was extended, first by adding site. The interest here was in whether site as a design variable moderated the impact of the treatment on outcomes (i.e., if there were differential effects of the intervention across sites).

Next, we conducted a set of supplementary analyses in which a variety of potential covariates were considered, some of which were meant to increase power (e.g., age, gender, ethnicity, free lunch status), whereas others were meant to account for potential instruction-related variability (e.g., additional school-provided instructional time). Although groups did not differ on many of these variables, their presence in models could reduce error variance if they were related to the dependent measures. Therefore, these relations were explored briefly. When they were included, they were evaluated as additional covariates,

rather than including all possible interaction terms because the mechanisms for how such variables would differentially affect the treatment groups is not clear. We were primarily interested in whether group differences remained similar with and without these additional variables.

Gender and ethnicity were generally unrelated to posttest measures, and where they were, they could be accounted for by other variables, so these were not further considered. Age was related to most posttest measures, including the standard scores from the WJ III and TOWRE, where it was used as a covariate as needed. Because of this, raw scores on these variables were also analyzed; results were substantively unchanged and are not further presented. In general, students who received free or reduced lunch performed less well on most measures relative to those who did not ( $p < .05$  on five of seven primary measures), so this variable was considered in the analyses.

Instruction-related variables included the amount of supplemental reading instruction received outside of the study, group size, and program coverage rate. Tutors' fidelity of implementation ratings were not included, because each tutor received a single average fidelity score that would be the same for all students taught by that tutor; thus an examination of the nested structure of the data would account for any variability in fidelity ratings. In the sample as a whole, the amount of additional supplemental reading instruction provided to students outside of the study was only weakly related to outcomes ( $Mdn r = .07$ ); these weak relations were negative, which may not be surprising in that it is likely that schools provided the most at-risk students with the most additional help. Intervention group size had a very small range (i.e., two to four), and did not differ on any outcome, so was not considered further. Program coverage rate (i.e., *Read Well* units covered per intervention session) was significantly related to six of the seven primary outcomes ( $Mdn r = .23$ ), so this variable was included in later models.

Finally, nesting was considered. We evaluated clustering in multiple ways, including by tutoring group, by intervention tutor, and by classroom reading teacher. We used SAS PROC MIXED to explore the effect of clustering in unconditional models as a function of tutor. The number of intervention group clusters was large (62), but the number of students per cluster was small (i.e., 2–4). Although the number of intervention tutors was somewhat low (12), the total number of students served by each tutor was generally high (range = 3–25), and this grouping seemed to most accurately reflect the impact of the tutor. There were 32 classroom teachers who each had from 1 to 15 intervention students in their classrooms, and this was the most ideal proportion of individuals and clusters, though perhaps not of the most direct relevance to this study. In general, clustering effects were strongest when defined according to instructional group ( $Mdn$  intraclass correlation coefficient [ICC] = 27%), then according to tutor ( $Mdn$  ICC = 13%), and least according to classroom teacher ( $Mdn$  ICC = 9%). The intervention tutor was used as the aggregate in the models described next, given its balance of cluster number and size.

## RESULTS

Pretest status is presented first. Then we present the primary results of the comparisons among treatment groups, considering only pretest performance. Inferential statistics are provided at an alpha level of  $p < .05$ . Next, supplemental analyses are presented, including site, instructional and noninstructional covariates, and nesting. Finally, we present the

**Table 2.** Pretest and posttest standard score performance on reading measures for treatment groups

Measure	N	Group	Pretest	8 Week	16 Week
			<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
WJ III Letter Word Identification	64	Concentrated	98.33 (10.25)	100.17 (10.65)	101.06 (10.82)
	65	Extended	95.74 (12.26)	96.82 (13.08)	99.20 (13.60)
	62	Distributed	96.60 (11.97)	96.90 (13.11)	99.71 (11.91)
WJ III Word Attack	63	Concentrated	106.57 (13.30)	106.13 (9.52)	105.41 (11.37)
	65	Extended	103.23 (14.10)	103.62 (13.15)	104.14 (14.49)
	61	Distributed	105.87 (13.38)	103.31 (13.57)	103.03 (14.69)
WJ III Spelling	64	Concentrated	—	98.33 (11.07)	98.84 (11.80)
	66	Extended	—	96.91 (12.62)	97.89 (13.03)
	62	Distributed	—	98.26 (12.31)	97.42 (12.40)
TOWRE Composite	64	Concentrated	—	91.72 (9.56)	92.11 (10.60)
	66	Extended	—	88.17 (10.96)	90.17 (11.63)
	62	Distributed	—	89.21 (10.99)	90.97 (12.26)
CMERS Passages	63	Concentrated	8.63 (3.55)	20.21 (9.95)	26.25 (15.09)
	64	Extended	7.88 (5.10)	17.40 (10.95)	25.06 (16.77)
	61	Distributed	9.34 (5.39)	19.89 (12.94)	26.50 (17.74)
WJ III Passage Comprehension	64	Concentrated	—	91.16 (9.49)	92.06 (9.04)
	66	Extended	—	86.92 (13.78)	89.32 (12.26)
	62	Distributed	—	87.50 (11.60)	90.81 (10.77)
GRADE Reading Comprehension	64	Concentrated	—	81.76 (8.66)	81.86 (9.07)
	65	Extended	—	82.31 (7.28)	83.29 (10.06)
	62	Distributed	—	83.56 (7.81)	83.89 (10.61)

*Note.* For Spelling, Letter Word Identification (LWID) was used as the pretest covariate; *N* is 65 for extended (at 8-week post only). For TOWRE, LWID was the pretest covariate; *N* is 65 for extended (at 8-week post only). For CMERS, there are 11 time points altogether (7 that spanned treatment). For Passage Comprehension, LWID was the pretest covariate; *N* is 64 for concentrated, 65 for extended, and 62 for distributed (at 8-week post only). For GRADE, LWID was the pretest covariate; *N* is 64 for concentrated, 65 for extended, and 61 for distributed (at 8-week post only). WJ III = Woodcock–Johnson Tests of Achievement III; TOWRE = Test of Word Reading Efficiency; CMERS = Continuous Monitoring of Early Reading Skills; GRADE = Group Reading Assessment and Diagnostic Evaluation; Concentrated = four times per week for 8 weeks; Extended = four times per week for 16 weeks; Distributed = two times per week for 16 weeks.

percentage of students who met criteria for adequate instructional response from each of the three groups.

### Pretest

Pretest and posttest means and standard deviations for the three treatment groups at the three primary time points are presented in Table 2. Note that pretest standard scores on WJ III Letter-Word Identification and Word Attack are relatively high, reflecting the fact that

only fluency criteria were applied to select the group and the relatively small number of items and lack of sensitivity of the WJ III for low-performing students at the beginning of Grade 1. Given that students were randomly assigned to groups, we expected that these groups would not differ from one another at pretest on any measure listed in Table 2, which they did not (all  $p > .05$ ).

### Research Question 1: Primary Results

Primary results are also contained in Table 2 as posttest performance with pretest as the primary covariate. We present results considering only pretest as a covariate.

*Decoding and Spelling.* On the WJ III Letter-Word Identification subtest (administered at pretest, 8-week posttest, and 16-week posttest), there was a significant effect of time,  $F(2, 187) = 32.68, p < .0001$ , indicating that all groups improved overall on this measure over the course of their first-grade year. However, there was no Time  $\times$  Treatment interaction,  $F(4, 374) = 1.50, p > .05$ , and the effect size was small ( $\eta^2$  is equal to  $1 - \lambda$ , or 3% in this case). For WJ III Word Attack (administered at pretest, 8-week posttest, and 16-week posttest), there was no significant effect of time,  $F(2, 185) < 1, p > .05$ , indicating that performance over the year was stable on this measure. There was also no Time  $\times$  Treatment interaction,  $F(4, 370) = 1.85, p > .05$ , and the effect size was small. For WJ III Spelling, Letter-Word Identification was used as the pretest covariate. At the 8-week posttest, there was no interaction of treatment group with the covariate,  $p > .05$ , and so this term was trimmed. There was a large pretest effect,  $F(1, 187) = 344.28, p < .0001$ , but no significant main effect for treatment,  $F(2, 187) < 1, p > .05$ . Results were similar at the 16-week posttest; pretest effect,  $F(1, 188) = 315.62, p < .0001$ ; treatment,  $F(2, 188) < 1, p > .05$ , with small effect sizes.

*Fluency.* The TOWRE was administered at the two posttest time points; Letter-Word Identification was used as the pretest covariate. At 8 weeks there was no interaction of treatment group with the covariate ( $p > .05$ ), and so this term was trimmed. There was a large pretest effect,  $F(1, 187) = 328.24, p < .0001$ , but no significant main effect for treatment,  $F(2, 187) = 1.18, p > .05$ . Results were highly similar at the 16-week posttest; pretest effect,  $F(1, 188) = 372.36, p < .0001$ , treatment,  $F(2, 188) < 1, p > .05$ , with small effect sizes. CMERS passages were administered at 11 time points, both prior to (for screening purposes) and during intervention. In a repeated measures context, there was a significant effect of time,  $F(10, 176) = 60.96, p < .0001$ , indicating that all groups showed strong improvement on this measure over the course of their first-grade year. However, there was no Time  $\times$  Treatment interaction,  $F(20, 352) = 1.08, p > .05$ , and effects were small at the posttest time points. Results were similar when only the time points during the intervention were considered.

*Comprehension.* Both comprehension measures were administered at the two posttest time points. For both measures, Letter-Word Identification was used as the pretest covariate. At the 8-week posttest, there was no interaction of treatment group with the covariate ( $p > .05$ ), and so this term was trimmed. There was a large pretest effect,  $F(1, 186) = 360.04, p < .0001$ , but no significant main effect for treatment,  $F(2, 186) = 2.04, p > .05$ . Results were similar at the 16-week posttest; effects were apparent for pretest,  $F(1, 188) = 332.49, p < .0001$ , though not for treatment,  $F(2, 188) < 1, p > .05$ . On the GRADE Passage

Comprehension subtest there was no interaction of treatment group with the covariate at the 8-week posttest ( $p > .05$ ), and so this term was trimmed. There was a significant effect for the covariate,  $F(1, 185) = 6.75, p < .02$ , but no significant main effect for treatment,  $F(2, 185) < 1, p > .05$ . Results were highly similar at the 16-week posttest; pretest,  $F(1, 187) = 5.80, p < .02$ ; treatment,  $F(2, 187) < 1, p > .05$ .

### Supplemental Analyses

*Effects of Site.* Site was systematically added to each of the aforementioned models. We were primarily interested in the extent to which site moderated intervention effects. For WJ III Letter Word Identification and Word Attack, site did not interact with treatment and/or time, though the smaller site did have significantly higher means, Letter Word Identification,  $F(3, 185) = 2.97, p < .04$ ; Word Attack,  $F(3, 183) = 5.22, p < .002$ . Results revealed no influence of site on WJ III Spelling or TOWRE. On CMERS, considering the seven time points between pre- and posttesting, there were no interactions or main effects of or with site and/or treatment. Site did not interact with treatment at either time point for either reading comprehension measure (WJ III or GRADE), although there was a main effect of site for the final time point of the WJ III,  $F(1, 187) = 6.86, p < .01$ , with the larger site showing significantly higher performance.

*Other Follow-Up Analyses.* As noted in the Analysis Plan, the only variables considered that were related to outcomes were age, free-lunch status, program coverage rate, and clustering. Other demographic and instructional variables were unrelated, or very weakly related, to outcomes, and so are not further reported. Age, free lunch status, program coverage rate, and site were considered together as covariates in additional models, with a specific focus on how these variables may have altered the treatment effects just presented. However, in a multivariate context, none of these covariates altered conclusions regarding treatment effects for any outcome variable. Finally, the effects of clustering were explored, but in no case did its inclusion change the conclusions regarding the effects of treatment.

### Research Question 2: Instructional Response

The adequacy of instructional response was assessed according to decoding, fluency, and comprehension benchmarks, based on performance at the 30th percentile (i.e., WJ III Basic Reading Skills Composite and Passage Comprehension standard scores of at least 93; ORF of at least 35 wcpm). As no percentile scores are available for CMERS ORF, we used the 30th percentile in the spring of first grade for Dynamic Indicators of Basic Early Literacy Skills ORF passages (Good, Wallin, Simmons, Kame'enui, & Kaminski, 2002). These benchmarks were selected to facilitate comparison with other first-grade reading intervention studies (e.g., Mathes et al., 2005). Presented in Table 3, the results indicate no significant group differences in proportion of students who met any of the responder criteria at the final time point: decoding,  $\chi^2(2, N = 192) = 1.09, p > .05$ ; passage fluency,  $\chi^2(2, N = 192) = 4.46, p > .05$ ; comprehension,  $\chi^2(2, N = 192) < 1, p > .05$ . Note that a percentage of students in each group met the Basic Skills criterion at pretest, when only ORF was a criterion for at-risk status.

**Table 3.** Percentage adequate responder status by treatment group

Criterion and Time	Treatment Group		
	Concentrated <sup>a</sup>	Extended <sup>b</sup>	Distributed <sup>c</sup>
Decoding			
Pretest	84%	68%	79%
8 weeks	89%	73%	69%
16 weeks	84%	79%	77%
Passage fluency			
Pretest	0%	0%	0%
8 weeks	11%	8%	10%
16 weeks	22%	17%	32%
Comprehension			
8 weeks	48%	39%	34%
16 weeks	47%	44%	52%

*Note.* For Decoding, criterion was a standard score of at least 93 the Basic Skills composite of the Woodcock–Johnson Tests of Achievement III (WJ–III), encompassing the Letter-Word Identification and Word Attack subtests. For Passage Fluency, criterion was an average score of at least 35 words correct per minute on two stories. For Comprehension, criterion was a standard score of at least 93 on the WJ–III Passage Comprehension subtest. Concentrated = four times per week for 8 weeks; Extended = four times per week for 16 weeks; Distributed = two times per week for 16 weeks.

<sup>a</sup>*n* = 64. <sup>b</sup>*n* = 66. <sup>c</sup>*n* = 62.

## DISCUSSION

The purpose of this study was to compare the effects on reading outcomes of delivering Tier 2 supplemental reading intervention to first-grade students at risk for reading difficulties on three different schedules. We hypothesized that students who received more extensive intervention would have better outcomes and a higher rate of intervention response than those on briefer schedules and that those on a distributed schedule would outperform those on a concentrated schedule. These hypotheses were not supported; there were no significant differences on any reading outcome or on the rates of adequate response to intervention between groups of students who received intervention on the three schedules. The addition of covariates related to site, demographics, nesting, and instructional variables did not change the conclusions.

### Intervention Duration

Contrary to the assumption that longer interventions are associated with higher gains than briefer interventions, this study found that first-grade students who were at risk for reading difficulties performed equally well following 16 and 32 hr of small-group intervention. Our findings contrast with those of Al Otaiba et al. (2005), who found in a randomized study that kindergarten students who received a year-long intervention four times per week had more robust outcomes than those who received the same intervention only two times per week. Differences in results for the two studies may stem from the facts that Al Otaiba and colleagues provided intervention throughout kindergarten and we began intervention

in January of first grade and that reading is highly emphasized in first grade and less emphasized in kindergarten.

A broader conception of treatment intensity—beyond time in intervention and group size—may be important. Researchers have suggested that characteristics such as the number of teacher–student interactions during each session (Warren, Fey, & Yoder, 2007), pacing both within and across lessons (i.e., program coverage rate; Mathes et al., 2011), and the level of active student engagement in instructional activities (Vaughn, Denton, & Fletcher, 2010), are related to intervention intensity. It may be that attention to these aspects of intensity is needed to accelerate the progress of at-risk readers. Future research that documents or manipulates these variables may provide insight into features of interventions and their implementation that are related to enhanced outcomes.

### **Instructional Response**

We found no significant group differences in rates of adequate instructional response related to intervention duration or scheduling. Moreover, these rates indicated that, on average, none of the groups appear to have received intervention with sufficient intensity to “powerfully accelerate” the development of broad reading proficiency for most students (Al Otaiba & Torgesen, 2007, p. 213–214). Although 77% to 83% of the students demonstrated adequate instructional response based on a decoding criterion, several met this criterion at pretest (as selection was based only on fluency). When considering the ORF benchmark of 35 wcpm, only 32% of students who received 29.5 hr of intervention and 20% of those who received 14.5 hr of intervention demonstrated adequate instructional response. On a reading comprehension measure, 44% to 52% students met the criterion of performance at the 30th percentile. Tier 2 interventions implemented for similar durations and at similar levels of intensity would be likely to leave large numbers of students in need of Tier 3 intervention.

Many students across all three groups may have required a more extensive intervention. In a systematic research synthesis, Wanzek and Vaughn (2007) found that providing reading interventions in small groups for at least 100 sessions is generally associated with medium to large effects sizes, particularly in kindergarten and first grade. Rates of adequate intervention response have been generally higher in our previous studies in which intervention began within the first 2 months of first grade and was provided daily for 30 to 40 min over 25 to 30 weeks (i.e., Denton et al., 2010; Mathes et al., 2005). In general, the proportions of adequate responders in the current study appear more similar to those of the typical practice comparison groups than the treatment groups in these more extensive first-grade intervention studies. For example, Mathes et al. (2005) provided first-grade at-risk readers with daily intervention in 40-min sessions for nearly 30 weeks, evaluating two comprehensive reading intervention programs. When adequate response was defined at the 30th percentile in the WJ III Basic Reading cluster, 93% of the students in the first program, 99% of those in the second, and 84% of a typical practice comparison group met the benchmark. When response was measured using an ORF benchmark, the response rates were lower but increased with more time in intervention. In a description of instructional response rates in the Mathes et al. (2005) study, Denton, & Mathes (2003) reported that, after 21 weeks of intervention, 37% of the students in one intervention and 46% in the other met the ORF benchmark of 35 wcpm, and after the full 30-week intervention 77% and 82% met the benchmark. A similar cumulative effect was observed in a second grade study by Vaughn, Linan-Thompson, & Hickman (2003), who found that students who began intervention

with the lowest fluency levels required more time to achieve adequate performance goals. The best approach may be to offer intervention throughout Grade 1 but evaluate student progress periodically and exit students who achieve benchmarks; however, Vaughn et al.'s study illustrated the need to continue to monitor progress of students who exit intervention after 10 or 20 weeks, as some will not continue to thrive with classroom instruction alone. The ideal Tier 2 system might allow such students to reenter intervention.

The failure to appreciably accelerate fluency development in the current study may be related to aspects of the intervention program and the way it was implemented, although our design did not allow us to directly evaluate the efficacy of the intervention or to separately evaluate its various components. Still, it is possible that our adaptations of *Read Well*, particularly the addition of vocabulary and comprehension lesson components, may have reduced the time and emphasis on phonemic awareness, decoding, and fluency required by these students, who were in the early stages of reading acquisition. It is also possible that the flexibility given the tutors to move through the program at a quick pace may have deprived students of the extended practice opportunities they needed to bring newly learned skills to automaticity. Our emphasis on program coverage may have compromised students' mastery of skills, although this could be considered a general disadvantage of brief Tier 2 interventions.

### Concentrated and Distributed Schedules

Contrary to our expectations, we found no differences in reading outcomes when students received 16 hr of intervention on more concentrated or distributed schedules, analogous to massed and distributed practice. This finding is aligned with that of Ukrainetz et al. (2009), who identified few differential outcomes for students who received phonemic awareness intervention on more concentrated and distributed schedules. Although laboratory-based experiments have reliably demonstrated the superiority of pacing instruction across several hours or days over presenting large amounts of content in a single session (e.g., Underwood, 1961), this principle may not apply to the scheduling of reading interventions in applied settings.

### Study Limitations

This study is limited by the absence of a no-treatment comparison group in the design. It is also important to recognize that this study directly contrasted only three of many potential intervention schedules. Our findings may not be generalizable to other reading intervention scheduling schemes, to intervention provided at other grade levels, to intervention provided during the first few months of first grade, or to intervention programs with a more limited focus on decoding and fluency. Finally, our inability to directly observe Tier 1 instruction and the additional reading instruction provided by schools outside of the research intervention limits the understanding of the context in which this study took place. In particular, we were unable to verify the extent to which classroom teachers differentiated instruction and implemented adaptations that were the subject of Tier 1 professional development.

### Implications for Research and Practice

Previous reading intervention research has established the potential of appreciably affecting student outcomes and learning trajectories by providing supplemental reading intervention

in the early grades (Fletcher et al., 2007). Although there is a continued need for experimental research evaluating programs and approaches used in reading instruction, there is also a need for research investigating the conditions under which certain interventions are most effective or which components of the interventions are essential and which are negotiable.

This study illustrated no differential outcomes when first-grade Tier 2 reading intervention is provided on more concentrated and distributed schedules. Additional study of these aspects of scheduling is warranted, especially as instructional approaches are designed for students at Tier 3 and those identified as having learning disabilities, who have intractable reading difficulties that are difficult to remediate. Designing interventions for these students may require investigations of dimensions of instruction and implementation that have not typically been the subjects of applied research.

Finally, there is a need for empirical research to guide practitioners in the implementation of interventions in an RTI context, including further research on scheduling and duration. One goal of RTI is to accelerate the progress of students who require support beyond quality classroom reading instruction so that they are able to read at average levels for their grades. It may be that first-grade reading interventions providing more than 32 hr of instruction would be more likely to accelerate the progress of at-risk readers than the intervention provided in this study. Alternately, it is possible that providing 32 hr of intervention similar to that in this study is sufficient to identify students who require more intensive intervention. If this is the goal of Tier 2, brief implementations may be sufficient, but it means more students will potentially need more complex and costly Tier 3 interventions. If the goal is to close the gap with average-performing peers at Tier 2, longer interventions may be warranted. Given the continued low reading performance of a significant proportion of students and the costs associated with providing supplemental intervention, such questions merit continued research.

## REFERENCES

- Al Otaiba, S., Schatschneider, C., & Silverman, E. (2005). Tutor-assisted intensive learning strategies in kindergarten: How much is enough? *Exceptionality, 13*, 195–208.
- Al Otaiba, S., & Torgesen, J. (2007). Effects from intensive standardized kindergarten and first-grade interventions for the prevention of reading difficulties. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention* (pp. 212–222). New York: Springer.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.
- Berkeley, S., Bender, W. N., Peaster, L. G., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, 42*, 85–95.
- Denton, C. A., Anthony, J. L., Parker, R., & Hasbrouck, J. E. (2004). The effects of two tutoring programs on the English reading development of Spanish-English bilingual students. *Elementary School Journal, 104*, 289–305.
- Denton, C. A., & Mathes, P. G. (2003). Intervention for Struggling Readers: Possibilities and challenges. In B. R. Foorman (Ed.), *Preventing and Remediating Reading Difficulties: Bringing Science to Scale* (pp. 229–251). Timonium, MD: York Press.
- Denton, C. A., Nimon, K., Mathes, P. G., Swanson, E. A., Kethley, C., Kurz, T., & Shih, M. (2010). The effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children, 76*, 394–416.

- Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? *Journal of Educational Psychology, 92*, 605–619.
- Fishman, E. J., Keller, L., & Atkinson, R. C. (1968). Massed versus distributed practice in computerized spelling drills. *Journal of Educational Psychology, 59*, 290–296.
- Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. New York: Guilford.
- Foorman, B. F., Fletcher, J. M., & Francis D. (2004). *Texas Primary Reading Inventory*. Austin: Texas Educational Agency and the University of Texas System.
- Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading*. Portsmouth, NH: Heinemann.
- Gansle, K. A., & Noell, G. H. (2007). The fundamental role of intervention implementation in assessing response to intervention. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention* (pp. 244–251). New York: Springer.
- Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., et al. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide* (NCEE 2009–4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/wwc/publications/practiceguides/>
- Glover, T.A. (2010). Key RTI service delivery components: Considerations for research-informed practice. In T. A. Glover & S. Vaughn (Eds.), *The promise of response to intervention: Evaluating current science and practice* (pp. 7–22). New York: Guilford.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.
- Good, R. H., Wallin, J., Simmons, D. C., Kame'enui, E. J., & Kaminski, R. A. (2002). *System-wide percentile ranks for DIBELS benchmark assessment* (Tech. Rep. No. 9). Eugene: University of Oregon.
- Grek, M. L., Mathes, P. G., & Torgesen, J. K. (2003). Similarities and differences between experienced teachers and trained paraprofessionals. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 267–296). Baltimore: Brookes.
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15*, 198–205.
- Harr-Robins, J. J., Shambaugh, L. S., & Parrish, T. (2009). *The status of state-level response to intervention policies and procedures in the West Region states and five other states* (Issues & Answers Report, REL 2009–No. 077). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved March 12, 2010, from <http://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=175>
- Hatcher, P. J., Hulme, C., Miles, J. N. V., Carroll, J. M., Hatcher, J., Gibbs, S., et al. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: A randomised controlled trial. *Journal of Child Psychology and Psychiatry, 47*, 820–827.
- Ikeda, M. J., Rahn-Blakeslee, A., Niebling, B. C., Gustafson, J. K., Allison, R., & Stumme, J. (2007). The Heartland Area Education Agency II problem-solving approach: An overview and lessons learned. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention* (pp. 255–268). New York: Springer.
- Kovaleski, J. F., & Black, L. (2010). Multi-tier service delivery: Current status and future directions. In T. A. Glover & S. Vaughn (Eds.), *The promise of response to intervention: Evaluating current science and practice* (pp. 23–56). New York: Guilford.
- Marston, D., Lau, M., & Muyskens, P. (2007). Implementation of the problem-solving model in the Minneapolis public schools. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention* (pp. 279–287). New York: Springer.

- Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly, 40*, 148–182.
- Mathes, P., Kethley, C., Nimon, K., Denton, C., Swanson, E., & Ware, P. (20XX). *The importance of measuring quality and quantity of implementation fidelity*. Manuscript under review.
- Mathes, P. G., & Torgesen, J. K. (2008). Continuous Progress Monitoring of Early Reading Skills (Version 3.0) [Web-based assessment in reading for Grades k-3]. Dallas, TX: Istation.
- Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19*, 107–122.
- Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 3*, 341–358.
- Spectrum K12 School Solutions. (2009). *Response to intervention (RTI) adoption survey 2009*. Retrieved June 15, 2009, from <http://www.spectrumk12.com>
- Sprick, M. M., Howard, L. M., & Fidanque, A. (1998). *Read well: Critical foundations in primary reading*. Longmont, CO: Sopris West.
- Torgesen, J. K., & Hudson, R. (2006). Reading fluency: Critical issues for struggling readers. In S. J. Samuels & A. Farstrup (Eds.), *Reading fluency: The forgotten dimension of reading success*. Newark, DE: International Reading Association. Retrieved from <http://www.fcrr.org/science/sciencePublicationsTorgesen.shtm>
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.
- Ukrainetz, T. A., Ross, C. L., & Harm, H. M. (2009). An investigation of treatment scheduling for phonemic awareness with kindergartners who are at risk for reading difficulties. *Language, Speech, and Hearing Services in Schools, 40*, 86–100.
- Underwood, B. J. (1961). Ten years of massed practice on distributed practice. *Psychological Review, 68*, 229–247.
- Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Paraeducator-supplemented instruction in structural analysis with text reading practice for second and third graders at risk for reading problems. *Remedial and Special Education, 27*, 365–378.
- Vadasy, P. F., Sanders, E. A., & Tudor, S. (2007). Effectiveness of paraeducator-supplemented individual instruction: Beyond basic decoding skills. *Journal of Learning Disabilities, 40*, 508–525.
- Vaughn, S., Denton, C. A., & Fletcher, J. F. (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychology in the Schools, 47*, 432–444. (PMCID: PMC2975106).
- Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69*(4), 391–409.
- Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*, 541–561.
- Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities, 41*, 126–142.
- Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13*, 70–77.
- Williams, K. T. (2001). *The Group Reading Assessment and Diagnostic Evaluation (GRADE): Teacher's scoring and interpretive manual*. Circle Pines, MN: American Guidance Service.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

## APPENDIX

*Read Well* Daily Lesson Plan

## Unit 16–Day 2

**DECODING PRACTICE:** 10 minutes1. Sound Review

Sound Cards: “You’re going to read the sounds about this fast (demonstrate). Read the sounds with me . . . I’m going to shuffle the cards. Read the sounds without me. [Student], now it’s your turn to read 2 cards.” Rotate easy cards in & out of practice. Keep all vowels in daily practice.

2. New Sound Practice

Decoding Sheet: “You’re going to say the sound that the 2 *o*’s make while you trace over the letters; I want you to write and say the sound: /oo/.” \*

3. Sounding Out Smoothly

Decoding Sheet: “Everyone, watch me say the sound that’s underlined, sound out the word and then read the word: /ea/, /heaeaeat/, *heat*.” Provide student practice. The word *moose* is the 1<sup>st</sup> word students read with a silent letter (*e*). Say “When you see a slash through a letter, it means the letter doesn’t say anything. Say the underlined part. Sound out your new word (/ooo/, /mmmmoooooosss/, *moose*). What is a moose (an animal)?” \*

4. Accuracy/Fluency Building

Decoding Sheets: “First say the underlined sound, then read each word.” (demonstrate; students practice) The word *scat* is the 1st word students read with the *sc-* blend. If they have difficulty, put the following practice exercise on the board: Write *at* & have students read *at*. Add a *c* & have students read *cat*. Add an *s* & have students read *scat*. \*

5. Tricky Words

Decoding Sheets & Tricky Word Cards: Tell students you will quietly count to 3 while they figure out each word. Say “Put your finger on the dot under the 1st word—(count silently)—1-2-3. Read the word. Move your finger to the next dot—1-2-3. Read the word.” Mix up flashcards for newer words & review 5 with the group & then call on individual students.

**VOCABULARY INTRODUCTION:** 3 minutes

1. Desert – (may be a review; taught in Unit 11, Day 3) Our first word is *desert*. Say *desert* (students repeat). A *desert* is a place that is very dry and sandy because it doesn’t get much rain. This is a cactus. (Show picture on p. 7) Cactus is a kind of plant that can live without much water, so cactus grows in the desert. If you went for a walk in the desert, what would you need to take with you? (If I went for a walk in the desert, I would take water, sun hat. . .) What is the word for *a place that is very dry because it doesn’t get much rain?* (desert)\*
2. Patient – Our next word is *patient*. Say *patient* (students repeat). If you are *patient*, you wait nicely and quietly for someone to come or for something to happen. It can be hard to be *patient* sometimes. It’s hard to be *patient* when you are waiting for Christmas to come. When is it hard for you to be *patient*? (It is hard to be patient when . . . accept

responses). If you *wait nicely and quietly for something or someone*, you are being what? (patient). Yes, you are being *patient*. \*

### **DAILY STORY READING & COMPREHENSION INSTRUCTION:** 10 minutes

#### **Story Books**

1. Reintroduce the storybook: Show students the unit title page again. Ask students if they remember what the animals in the story are looking for (the moon).
2. Duet Story: In the Night Sky – Read the title of Chapter 3 to students. The teacher reads the small text and students read the large text.
  - TELL vocabulary word *sea* on p. 6: *The sea is the same as the ocean. It is where boats go out on the water.* Do you think the moon was in the sea? (no)
  - TELL vocabulary word *forest* on p. 6: *The forest is a place with a lot of trees.* (Point to picture.) Do you think the moon was in the forest? (no)
  - TELL vocabulary word *sink* on p. 7: *To sink means to go down into water or something else. A boat can sink.* Do you think the moon could sink into the sand in the desert? (Point to the picture when you say the word *desert*.) (no)
3. Solo Story: In a Hat? – Ask students to read the title of the story. Students read all text.
  - Before they read, TELL the vocabulary word *shack* on p. 3. *A shack is a little, old building* (Point to picture).
  - Reread Solo Story as time allows
    - First Solo Read: Choral Reading
    - Second Solo Read: Beat the Clock Reading (Fluency)

#### **Oral Comprehension Discussion\*\***

If this group started on Unit 1: List 2–4 of the main events

“Remember from the last lesson that every story or movie has main things that happen that make the story interesting. What are some of the important things that happened in this story?”

If this group started on Unit 10:

“Remember that every story or movie has some sort of problem that the characters are trying to solve or find the answer to. What is the problem in our story?”

### **UNIT ASSESSMENTS:** 6 minutes

Unit 16 Decoding Assessment

### **INDEPENDENT FLUENCY/WRITING PRACTICE**

Paired or individual reading of new unit stories or previous unit stories

Writing Practice: oo, has, do, would, into

### **DAILY HOMEWORK** – Reprint of Story In a Hat

\*Portion of script removed to conserve journal space. \*\*Placement tests were used, so groups started intervention on different lessons. Progression of instruction for all groups was character-setting-problem-events-solution, but groups would be at a different point in this process depending on where they started.