Published in final edited form as:
J Res Educ Eff. 2014 ; 7(3): 250–267. doi:10.1080/19345747.2014.906692.

Long Term Effects of First Grade Multi-Tier Intervention

Stephanie Al Otaiba, Ph.D. [Professor],
Department of Teaching and Learning Annette Caldwell Simmons School of Education and Human Development Southern Methodist University PO Box 750455 Dallas, TX 75275-0455 214-768-1339 salotaiba@smu.edu

Young-Suk Kim, Ed.D. [Associate Professor],
School of Teacher Education Florida Center for Reading Research Stone Building G129 Florida State University Tallahassee, Florida 32306-4459 850-644-0370 ykim@fcrr.org

Jeanne Wanzek, Ph.D. [Associate Professor],
School of Teacher Education Florida Center for Reading Research Psychology C234U Florida State University 1107 W. Call St. P.O. Box 306-4304 Tallahassee, FL 32306 850-644-9080 jwanzek@fcrr.org

Yaacov Petscher, Ph.D. [Director of Research], and
Florida Center for Reading Research 2010 Levy Ave Suite 100 Tallahassee, Florida, 32310 850-645-8963 Ypetscher@fcrr.org ℗

Richard K. Wagner [Robert O. Lawton Distinguished Professor of Psychology W Russell and Eugenia Morcom Chair Associate Director]
Florida Center for Reading Research Department of Psychology, Room A205 Florida State University 1107 West Call Street PO BOX 3064301 Tallahassee, FL 32306-4301 rkwagner@psy.fsu.edu

Abstract

The purpose of this study was to compare the long term effects of two first grade RTI models (Dynamic and Typical RTI) on the reading performance of students in second and third grade. Participants included 419 first grade students (352 in second grade and 278 in third grade after attrition). Students were classified based on first grade screeners as at-risk or not at-risk and then based on their response to intervention (no risk [NR], relative easy to remediate [ER] and requiring sustained remediation [SR]). Students in the Dynamic RTI condition had higher reading comprehension scores at the end of third grade. At the end of second grade, ER and SR students had lower reading scores than NR students. At the end of third grade, there were no differences in reading skills between ER and NR students, but SR students had lower scores than NR students. ER students in the Dynamic RTI condition had higher reading scores at the end of second grade than those in the Typical RTI condition. Limitations and directions for future research are discussed.

Keywords

Response to intervention; reading; learning disabilities; longitudinal

The reauthorization of the Individuals with Disabilities Improvement Act in 2004 allowed local education agencies to provide multi-tiered models of early reading intervention through Response to Intervention (RTI). Researchers and policy makers were optimistic that RTI would allow students to receive increasingly intensive tiers of support without requiring students to wait to fail, which was a major criticism of the use of the IQ-achievement discrepancy model for identifying students with reading disabilities (e.g., Bradley, Danielson, & Hallahan, 2002; Fletcher et al., 2004). In addition, it was hoped that the prevalence of reading disabilities and of students with reading difficulties would be reduced by ensuring students received high quality classroom reading instruction (e.g., Foorman, Francis, Fletcher, Schatschneider & Mehta, 1998). Furthermore, there was converging evidence from studies, many of which were funded through the National Institute of Child Health and Human Development, that explicit and systematic early reading interventions were effective (Lyon & Chabra, 1996; Mathes et al., 2005; National Reading Panel, 2000; Simmons, Kamee'nui, Stoolmiller, Coyne & Harn, 2003; Torgesen, Wagner, Rashotte , Rose, Lindamood, Conway et al, 1999; Authors, 2007). Essentially, proponents of RTI expected the process to help prevent reading difficulties due to poor instruction or experiential differences by ensuring all students had access to research-based, early reading instruction and intervention. (e.g.,Vellutino et al, 1996; Vellutino, Scanlon, Small & Fanuele, 2006). In the process, students with true cognitive-based reading disability could be identified. However, there is limited knowledge of the Tier 3 interventions for students with the most significant difficulties and disabilities; fewer still which have examined response to intervention longitudinally (Lam & McMaster, in press). Such knowledge could guide the next generation of RTI and special education research, particularly in terms of intervention design for students with persistent reading problems.

Longitudinal Multi-tier RTI Research

As many as nine experimental or quasi-experimental RTI implementation studies have included Tier 3 intervention (Authors, in press; Beach & O'Connor, 2013; Denton, Fletcher, Anthony & Francis 2006; Denton, Tolar, Fletcher, Barth, Vaughn & Francis, 2013; Gilbert et al., 2013; O'Connor, Harty, & Fulmer, 2005; Vaughn, Wanzek, Linan-Thompson & Murray, 2007; Vaughn, Wanzek, Murray, Scammacca, Linan-Thompson, & Woodruff, 2009; Vellutino, Scanlon, Zhang & Schatschneider, 2008). Notably, these studies consistently included at least three of the five core components deemed essential for RTI implementation according to Gersten et al's (2009) *Practice Guide for Response to Intervention* (universal screening, progress monitoring, increasingly intensive tiers of intervention, and fidelity of implementation). However, with regard to the component of a high quality core reading program, the effectiveness or quality of Tier 1 instruction has not been consistently observed; thus it is not clear that the field has yet established that students provided a Tier 2 or 3 intervention had also received high quality core or Tier 1 instruction (Hill, King, Lemons, & Partanen, 2012). Furthermore only a few studies have provided increasingly intensive layers of intervention *within* a study year and/or have allowed students to fluidly move up *or* down a tier as may be necessary in school-implemented RTI models (e.g., Authors, Beach & O'Connor; O'Connor et al).

In a recent literature review of response to RTI, Lam and McMaster (in press) extended prior syntheses describing responsiveness to multi-tier interventions (e.g., Authors; Nelson, Benner & Gonzalez, 2003; Tran, Sanchez, Arellano,& Swanson, 2011) by reporting that students' initial word identification, alphabetic principle, phonemic awareness, and fluency consistently predicted their responsiveness. Their findings appear to corroborate Vellutino et al's (2006) hypothesis that inadequate responders demonstrate a pattern along a continuum of severity on a variety of reading-related variables. Lam and McMaster echoed growing concern from the field that students with the weakest initial skills might be required to remain in Tier 2 for too long, making RTI another potential wait to fail system (Vaughn, Denton, & Fletcher, 2010).

The few research teams who have examined Tier 3 interventions to date have reported mixed findings with large individual differences in responsiveness after providing more intensive interventions to students who had previously demonstrated inadequate response to Tier 1 and 2. Such interventions have included a substantial focus on phonics, fluency, and comprehension. In an early and seminal study, O'Connor and colleagues (O'Connor, 2000; O'Connor et al., et al. 2005) provided a three tier model from kindergarten through first grade and tracked students through third grade. This complex quasi-experimental study provided preliminary evidence that struggling readers who received Tier 2 or Tier 3 interventions performed better than struggling readers who received only Tier 1. O' Connor's RTI model allowed students to fluidly move up or down the intensity spectrum as needed and reported that, in general, the percentage of students at-risk decreased when additional tiers were provided. These tiers increased in intensity by providing more time in intervention, as well as smaller group size. A majority of students who received Tier 2 intervention (but did not need Tier 3 and thus were more easily remediated) achieved grade level standards. However, only 4 of the 10 who needed Tier 3 could read on grade level by the end of third grade; thus 60% were inadequate responders to Tier 3.

Denton and colleagues conducted two studies examining the effect of Tier 3 for primary grade students who had demonstrated inadequate response to well-implemented Tier 1 and Tier 2 in first grade (Denton, Fletcher, Anthony, & Francis, 2006; Denton et al., 2013). In the first study, Denton and colleagues located students who had not responded to an earlier first grade intervention (Mathes et al., 2005) and added additional students who were in second and third grade with similar reading scores; a total of 27 inadequate responders received intensive Tier 3 intervention (half had participated in the prior study and received Tier 1 or Tier 2 but half had not). However, more than half (n = 15) demonstrated little or no growth, defined as gains of less than 0.5 standard scores on a measure of basic reading skills. Nine of the 27 were third graders and seven already had already been identified for special education. In the second study, students in second grade (and a small number repeating first grade) who had demonstrated inadequate first grade response were randomly assigned to receive an individualized intervention for Tier 3 ($N = 47$) or to receive school-delivered intervention ($N = 25$). The authors reported that students in the individualized intervention showed significantly better improvement than school controls on word identification, phonemic decoding, word reading fluency, and on sentence-and paragraph reading comprehension. There were large individual differences in response with 72% of students performing within the average range (above a 90 standard score) for word reading

and word attack. However, students ended the Tier 3 intervention with impaired oral reading fluency (on average they read 51 correct words per minute, or below the 20th percentile; notably this score represents the 50th percentile for the end of first grade) and with impaired comprehension (only 36% achieved a standard score of 90, and the mean standardized scores hovered near the 20th percentile as well).

Authors (2008) also examined more intensive intervention for students with inadequate response to Tier 2. In two studies, they examined student response to (a) continued single dose of intervention (daily, 30 min sessions), (b) a double dose of intervention (daily, 60 min sessions), or (c) no research-provided intervention. More students in the interventions accelerated learning than in the no intervention group. However, response to the single-dose and double-dose interventions was similar. Vaughn et al. (2009) also provided Tier 3 second grade intervention to students who had not responded to Tier 1 and 2 in first grade. They reported significant differences in gain scores compared to controls on untimed word reading and reading comprehension. Students in the intensive intervention who had higher initial ORF scores made the most progress in word reading and comprehension.

Despite some early successes with these Tier 3 interventions, across the studies there appeared to be a pattern of stronger response on word reading skills than on either reading fluency or reading comprehension. The one exception was that Denton et al. (2013) found that students showed more progress in word reading and comprehension than in reading fluency. Furthermore, there are some troubling findings from longitudinal follow ups for students participating in RTI studies that include Tier 3 intervention. For example, Vellutino et al. (2006) reported that most poor readers who were difficult to remediate ended first grade with basic reading skills within the average range ($M = 94.63$ and $M = 93.32$ for word identification and word attack). It was encouraging that from first to third grade, their mean reading comprehension standard scores actually increased from 79.16 to 88.47. However, basic reading scores decreased in second and third grade; by third grade their mean scores had dipped to 79.16 and 79.95 for word identification and word attack, respectively. In fact, nearly a third of these students who had been the most difficult to remediate no longer scored within the average range for basic reading skills (with standard scores of 90 or below). In contrast, students who were easier to remediate initially did continue to read within the average range in third grade, suggesting that once intervention ended, they were able to benefit from typical classroom instruction.

Gilbert et al. (2013) also reported an increase in the proportion of students who could not read in the average range from the end of the study intervention (first grade) at the end of second and third grade. They reported 60% of first grade students with previous insufficient response who were provided continued Tier 2 read below the 30th percentile (defined as reading below this cut point on timed and untimed real and pseudoword reading) by the end of third grade. Of the students who received Tier 3 intervention in first grade, 46% fell below the 30%th percentile by the end of third grade.

The findings from these two studies suggest the importance of longitudinal follow-up for students with the most severe reading difficulties to better inform intervention practices. Yet, given this paucity of longitudinal research, there is a need to confirm whether early RTI

prevents future problems. There is also a need to examine whether adequate initial response leads to normalized reading outcomes for over time. In addition, there is a need to consider what interventions might be designed for students with persistent reading difficulty despite having participated in effective first grade multi-tier interventions. Recently, we (Authors) compared the short term effects of two models of RTI implementation in first grade with the intention to follow the students through elementary school. Details on the models implemented are described in more detail in the methods section but briefly, one model, Dynamic, fast-tracked students to Tier 2 or Tier 3 interventions according to initial need and the other used a more traditional two-stage process that required students to begin in Tier 1 and only receive intervention if they did not respond to each successive tier. Interventions used in both models were identical; thus the only difference was when intervention began. Reading assessments included letter-sound, word, and passage reading, and a teacher-report of the severity of student reading difficulties. Using multi-level modeling, the intent-to-treat analysis indicated an overall effect favoring the Dynamic RTI condition ($d = .36$). Moreover, the growth curve analyses demonstrated that students in Dynamic RTI showed an immediate score advantage, and that effects accumulated across the year. In addition, students in the Dynamic condition who received Tier 2 and Tier 3 ended the study with significantly higher reading performance than students in the Typical condition. Across the two conditions, only 12.82 % of students who received Tier 2 and 3 did not achieve basic reading skills (defined as a standard score of 90 or more) (Authors).

Purpose of the Present Study

The purpose of this study is to report and analyze the longitudinal outcomes of students who participated in RTI implementation in first grade. Specifically, we categorized student response to intervention after the first grade RTI implementation into three groups: (a) students who were not at risk “NR” (who therefore received only Tier 1 and did not receive Tier 2 or Tier 3 intervention), (b) students who were easy to remediate “ER” (students received intervention, responded, and were subsequently moved to less intensive interventions in the model), or (c) students who were required sustained or increasingly intensive remediation throughout the year “SR”.

Thus we had two specific aims for the present study. The first aim was to examine end of second and third grade outcomes for these three groups in order to learn whether there were differences in outcomes based on the type of RTI model implemented. The second aim was to report the proportion of students within these risk categories who were performing below a standard score of 90 at the end of second grade and third grade on a variety of timed and untimed reading measures.

Methods

Participants

Data are from 419 students (236 boys) who participated in the initial RTI study. These students were in schools identified in 2008, with the help of the District Reading Specialist in a mid-size southeastern school district. One school opted out of the follow up, thus the present study includes students from six schools and 25 first grade general education

classrooms. By grade three, these students spread into 44 classrooms. The majority of students were African Americans (57%) and Whites (32%). Approximately 68% of the students were eligible for free and reduced lunch. The majority of schools had more than 50% of students receiving free and reduced price lunch (FARL ranged from 42.8% to 89.9%), but, consistent with district demographics, few students were Limited English Proficient.

Of the 419 students, 352 students remained at the end of second grade (i.e., 16% attrition) and 278 students remained at the end of third grade (34% attrition). Little's test (Little, 1988) of missing data completely at random showed that missingness were not completely at random $\chi^2(8) = 26.75$, $p = .001$ in grade 2 and $\chi^2(10) = 27.05$, $p = .003$ in grade 3. Therefore, multiple imputation was conducted to correct for potential bias in parameter estimation. Imputation was conducted at the student level using SAS PROC MI procedure for the four reading outcome measures in second and third grade. Fifty imputations were conducted using a Markov Chain Monte Carlo estimation.

A total of 208 students were in the Dynamic condition and 211 in the Typical condition. Students in the NR group were those who stayed in Tier 1 throughout the year ($n = 262$). The ER students were those who were reclassified to lower tiers during the year due to their positive response to intervention ($n = 31$), including those who moved from Tier 3 to 2 or from Tier 2 to 1. Students in the sustained remediation (SR) were those who stayed in the same tier or were reclassified to higher tiers ($n = 126$), including those who stayed in Tiers 2 and 3 throughout the year and those who moved from Tier 1 to 2, from Tier 2 to 3, and from Tier 1 to 3. Table 1 describes demographics and scores on the reading measures by group.

Prior First Grade Research Design: Understanding the Instructional Setting and the Three Tier Intervention and Conditions

A thorough description of the first grade study is provided in (Authors, in press). What follows is a brief description of Tier 1and documentation of the quality of Tier 1 instruction, a description of the two RTI models, Tier 2 and 3 interventions and the fidelity of implementation, and a description of procedures for screening and assignment to condition and tier.

Tier 1—For Tier 1, all teachers used *Open Court* as the core reading program (Bereiter et al., 2002) and all participated in a one day workshop the summer prior to the study to help them learn about RTI generally, and to learn about our experiment design and our reading interventions. Tier 1 reading instruction was videotaped and the effectiveness of implementation was rated with a low-inference observational instrument used in prior research (Authors, 2011; Authors, 2011) and adapted from Haager, Gersten, Baker, and Graves (2003). Averaged across both two observations, teachers' overall reading instruction was rated as effective ($M = 1.88$, $SD = .36$); range (1 - 2.69). Our scale ranged from 0, indicating content was not observed and, 1, 2, 3 respectively, for *not effective, effective*, and *highly effective*. Researchers established inter-rater reliability (Cohen's kappa of.975 and . 972 for fall and winter respectively).

Dynamic and Typical conditions and Tier 2 and 3 interventions—As previously noted, the two RTI conditions were designed to be identical in the following ways: (1) the standard protocols for intervention at Tier 2 and 3 were identical across conditions interventions and were closely aligned with Tier 1 reading and language arts instruction and therefore staff did not modify instruction significantly; (2) well-trained project staff provided the Tier 2 and 3 interventions; (3) students could move up to a more intensive tier if needed and could move down to a less intensive tier when they were successful for two screening periods. Thus, the only difference between conditions was *when* students were provided supplemental intervention sessions.

In the first RTI condition, Typical RTI which was similar to two-stage screening models in that all students began in Tier 1 and following a second screening (8 weeks later) were eligible for Tier 2 only if they had not responded to Tier 1. Subsequently, students who demonstrated insufficient response to Tier 2 intervention were provided a more intensive Tier 3 intervention during the third session. By contrast, in Dynamic RTI, students were provided Tier 2 or Tier 3 according to their initial screening or to criteria at subsequent screenings.

For Tier 2, well-trained interventionists (advanced graduate students and certified teachers) provided two, 30 min weekly sessions in groups of 4-7 students. For Tier 2, code-focused activities were drawn from the first grade Open Court *Imagine It*! (Berieter et al., 2002) series and the Florida Center for Reading Research K-3 Center Activities (www.fcrr.org) and included phonological awareness and letter sound skills, decoding and sight word instruction, and fluency training. For Tier 3, interventionists provided four, 45 min weekly sessions in smaller groups of 1-3 students using the standard protocol for *Early Interventions in Reading* (*EIR*; Mathes, Torgesen, Wahl, Menchetti, & Grek, 1999) that includes phonemic awareness, alphabetics and phonics, and fluency (roughly 30 min).

Across both Tier 2 and Tier 3, interventionists also provided meaning-focused instruction for about 10-15 min per day. In the first eight weeks, they read aloud high interest trade books using dialogic reading techniques (e.g., Lonigan & Whitehurst, 1998). In the second 8 weeks, students read decodable books to practice, in the third 8 weeks students read decodable books written to emphasize the sequencing text structure (i.e., first, next, and last) and retold the story.

For fidelity purposes, interventionists videotaped intervention sessions every eight weeks. These videotapes were scored by a master coder and a senior staff member using a fidelity checklist and their inter-rater reliability was high (98.1%). Fidelity ratings for the tutors ranged from .77 to .98 (M = .89).

Procedures for screening and assignment to condition and tiers—In September of first grade, we asked teachers to complete a rating of the severity of students' reading difficulties relative to classmates (Speece et al., 2011). Then trained graduate students administered four screening measures which we used in addition to the teacher checklist to determine initial eligibility for Tiers 1-3. These included both subtests of the Test of Word Reading Efficiency (described below in the measures section) and two, one minute

curriculum based measures. These CBM tasks have been used in prior RTI studies; the first was AIMSWEB Letter Sound Fluency (Shinn & Shinn, 2004). In this task, children are presented an array of 10 rows of 10 lower case letters per line and are asked to name as many of the sounds the letters make as quickly as they can. Testing is discontinued if no correct sounds occur in the first row. Raw scores are reported and alternate form reliability is .90. The second was the Word Identification Fluency (WIF) task (Fuchs, Fuchs, & Compton, 2004). In this task, students read from an array of 50 first grade sight words (randomly selected from the Dolch word list of 100 frequent words). Raw scores are reported and alternate form reliability was reported to be .97 (Fuchs et al., 2004). We then created z-scores (from the raw scores on these four measures) and identified students who scored below the 40th percentile, representing low average performance, within their schools. Within the larger study, schools had variable socioeconomic status levels and reading scores, so we used school/local norms or cut points to determine eligibility, and we excluded students who scored above a standard score of 95 on word identification and passage comprehension.

Next to determine initial eligibility for tiers, students whose teachers reported they had severe reading difficulties *and* who scored below the 40th percentile at the school level for all four screeners were considered as initially eligible for Tier 3. Students whose teachers reported they had severe reading difficulties *or* who scored below the 40th percentile at the school level for three out of four screeners were considered initially eligible for Tier 2.

Then to assign students to condition, which by design would determine when they would begin intervention and at what tier, students were rank ordered within classrooms (on a sum of the z-scored screening results) and pairs were randomly assigned to the Dynamic condition which allowed them to enter their eligible tier immediately or to the Typical condition, which required them to complete 8 weeks in Tier 1 and then move through Tier 2 and 3 only if they demonstrated inadequate response at subsequent screeners. Thus, after the first 8-week session, which corresponded to the report card period, all students (in Tiers 1-3) were again screened to re-determine local norms. Students who remained below the 40th percentile on three out of four measures and who also demonstrated slopes of growth less than the mean for the entire sample moved to a more intensive tier in the next eight week session (e.g., from Tier 2 to Tier 3 in Dynamic or from Tier 1 to Tier 2 for Typical). When students were easily remediated (i.e., they scored above the 40th percentile and demonstrated slopes of growth at or above the mean) in a tier for two consecutive eight week periods, they were exited to a less intensive tier.

Measures and Data Collection Procedures

All data used in the present study were collected by a highly trained research team. Students were individually assessed in a quiet are within their school.

Word reading accuracy—Student's word reading accuracy was assessed using the Letter Word Identification of the Woodcock-Johnson III Test of Achievement (WJ-III; Woodcock et al., 2001). In this task, the student is asked to read aloud words of increasing difficulty. Reliabilities range from .94 to. 98 for ages 6 to 9.

Word reading efficiency—The Sight Word Efficiency and Phonemic Decoding Efficiency subtests of the Test of Word Reading Efficiency (TOWRE, Torgesen, Wagner, & Rashotte, 1997) were used in grades 1 and 2. In third grade, TOWRE-2nd edition (Torgesen, Wagner, & Rashotte, 2012) was used. In these tasks, students are given 45 seconds to read a list of sight words and a list of pseudo-words that increase in difficulty. Alternate forms and test-retest reliability coefficients for each subtest exceed .90.

Oral reading fluency—In Grades 1 and 2, the Oral Reading Fluency subtest of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2001) was used. It is a brief 1-min measure where students are asked to read a grade-level passage and the number of correct words per minute is their score. Three passages are given for the assessment and for this study, mean raw scores were be analyzed. Test-retest reliability for elementary students is .92. In third grade, the Oral Reading Fluency subset of Wechsler Individual Achievement Test – Third Edition (WIAT-3; Wechsler, 2009) was used. The task administration is essentially the same as the DIBELS oral reading fluency task. Students are asked to read aloud two previously unseen passages. The time taken to read each passage and number of misread words are recorded. The score is the number of words read accurately per minute. Reliability is .94. Standard scores are available and reported in Table 1.

Reading comprehension—The Passage Comprehension subtest of WJ-III (Woodcock et al., 2001) was used. This is an oral cloze task in which the student is given sentences and short passages and is asked to provide correct answer in blanks. Reliabilities range from .91 to. 96 for ages 6 to 9.

Results

Descriptive Statistics

Table 1 presents students' performances on reading measures in the fall and spring of grade 1, and spring of grades 2 and 3 for the three groups. As expected, students in the NR group had higher performances on all the reading tasks across the year. Children in the NR group performed in the average range although in some measures their standard scores were in the high average range (e.g., 112.77 in TOWRE at the end of second grade). Average standard scores for students in ER group varied from 82.90 to 108.00 across measures and times. Performance levels for students in SR group varied from 83.47 to 99.17.

Research Aim 1: Long Term Effects of the Two RTI Interventions on the No Risk (NR), Easily Remediated (ER, and Sustained Remediation (SR) Groups

Multi-level analysis (students nested in classrooms) employing SAS 9.4 PROC MIXED procedure and PROC MIANALYZE was conducted using imputed data for the four reading outcomes, WJ Letter Word Identification, TOWRE, oral reading fluency, and WJ Passage Comprehension. All the analyses were conducted using standard scores except for oral reading fluency in second grade for which we used raw scores due to absence of standard scores. When unconditional models were fitted, intraclass correlations at the end of second grade were as follows: .12 in Letter Word Identification, .11 in TOWRE, .14 in oral reading

fluency, and .05 in Passage Comprehension. Intraclass correlations at the end of third grade were as follows: .25 in Letter Word Identification, .19 in TOWRE, .34 in oral reading fluency, and .31 in Passage Comprehension. In other words, approximately 5 to 34% of variations in the four reading measures were attributable to classroom differences in second and third grade.

In order to examine whether the first grade RTI models had a long-term effect on students reading skills in second and third grade and whether the effect varied as a function of students' group (i.e., NR, ER, and SR) in first grade, dummy codes for first grade intervention condition (i.e., Dynamic vs. Typical), and interaction terms for the interaction between intervention condition and students' group status were created. Students' demographic variables such as gender, free and reduced lunch status, and racial backgrounds were included as control variables. In addition, children's performance in each of the literacy scores in the fall of grade 1 was included as a control variable (i.e., for the Letter Word Identification outcome in grades 2 and 3, children's fall grade 1 on the Letter Word Identification was included as a control variable). Finally, given multiple comparisons, statistical significances were examined applying Benjamini-Hochberg (1995) corrections. For guidelines for the Benjamini-Horhberg corrections and effect size calculations, see What Works Clearinghouse recommendations (http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19).

Results for second grade are presented in Table 2. For the Letter Word Identification outcome in second grade, students in the ER and SR groups had statistically lower scores on TOWRE and oral reading fluency compared to the students in the NR group (*ps* < .01). In addition, SR children also had statistically lower scores on the Letter Word Identification (*ps* < .01). Interestingly, the interaction between the Dynamic condition and ER was significant and positive such that ER students in the Dynamic condition had higher scores on all the four reading outcomes in second grade (*ps* ≤ .02). To illustrate this interaction effect for the Letter Word Identification outcome, the expected standard scores were as follows: 114 for ER children in the dynamic condition, 100 for ER children in the traditional condition, 104 for SR children in the dynamic and traditional conditions. Effect sizes for the interaction terms in the Letter Word Identification, TOWRE, oral reading fluency, and Passage Comprehension were 1.35, 1.75, 1.66, and 1.10, respectively. The interaction between the Dynamic condition and SR was not statistically significant on any of the outcomes (*ps* ≥ .55).

Results in third grade reading scores are presented in Table 3. Note that by third grade, due to attrition, there were no ER students who had been in the Typical condition, the interaction between ER and Dynamic condition could not be examined. ER students have statistically higher scores than NR students (*ps* ≤ .03) in all the reading outcomes (*ps* < .01) except Passage Comprehension whereas SR students had significantly lower scores than students in the NR group in these three reading outcomes. The effect sizes for the ER students were .58, .78, and .55 in the Letter Word Identification, TOWRE, and oral reading fluency, respectively. Effect sizes for the SR students were −.44, −.81, and −.47 in the Letter Word Identification, TOWRE, and oral reading fluency, respectively. The interaction between intervention condition and SR was not statistically significant (*ps* ≥ .20).

Research Aim 2: Proportion of NR, ER, and SR Students with Reading Standard Scores below 90 in Second and Third Grade

In order to examine proportions of students with low and below average performances on reading tasks, in keeping with criteria used in prior research (e.g., Denton et al., 2013) we identified students with a standard score of below 90 (25%ile) at the end of second and third grade. Note that due to the lack of normative information for DIBELS oral reading fluency, which was used in second grade, we used a score (i.e., 61) corresponding to the 25th percentile according to the Hasbrouk and Tindal's (2006) norming table (similar to Denton et al.). The results for the frequencies and percentages are presented in Table 2. Overall, few students were identified to be poor in all the measures and this appears to be attributable to the fact that few students had scores below 90 on the WJ Letter Word Identification task. In general, there was a pattern of more students below a 90 for timed tasks and for comprehension than for word identification. Not surprisingly, a lower percentage of NR students were below 90 than those in the ER and SR groups.

However, among students in the NR group between 1% and 11.8% had standard scores below 90 at the end of second and third grade. For example, on the WJ Passage Comprehension, the percentage of students scoring below 90 was 11.8% at the end of second grade and approximately 8% at the end of third grade. For students in the ER group, the percentage of students below 90 varied from 3% to 13% on the various measures. For students in the SR group, this percentage ranged from 7.1% to 42.1% and over 30% of students had scores below 90 in the TOWRE, oral reading fluency, and WJ Passage Comprehension.

Discussion

The present study was designed as a longitudinal follow up to our prior randomized control trial investigating the efficacy of two models of RTI (Authors, in press). Given the relatively limited longitudinal research on response to multi-tier interventions, the first research aim was to compare the long-term effects of participation in two first grade multi-tier intervention models on second and third grade reading scores and to learn whether the long term effects varied as a function of students' first grade responsiveness (i.e., NR, ER, and SR). Notably at the start of first grade, the standard scores across reading measures for the ER and SR students were similar and mostly in the low average range; by contrast the NR group had standard scores of 100 and above. Students in the SR group had significantly lower scores at the end of second and third grade than did students in the ER and NR groups. The results of the multi-level analyses extend our earlier first grade randomized control trial and indicate that students in the Dynamic condition significantly outperformed students in the Typical condition on word identification at the end of second grade, but that this difference was no longer significant at the end of third grade. The significant interaction demonstrated that the Dynamic condition resulted in better long-term outcomes (until the end of second grade) on all four reading measures; unfortunately since there were no students in the ER condition in the Typical condition in third grade, we could not compare these groups at the end of third grade. Further research is needed to replicate these results but a potentially important implication is that it is not necessary to wait for students to not

respond and that 8 week intervals for moving up or down tiers in terms of intensity was feasible and more effective for improving children's reading skills, particularly for children who were initially relatively easy to remediate. It appears that when students received interventions that were matched in intensity to the severity of their needs, they performed better in the long-term. Clearly another important implication is that most children identified as needing additional intervention will continue to need this for a sustained time, and many will require even more intensive intervention.

Our second aim was to describe the proportions of students scoring below a standard score of 90 at the end of second and third grade. We used scores at the 25%ile and below (standard score of 90) to allow us to directly compare our results with other research teams using this cutpoint (e.g., Denton et al., Fletcher et al); We found that a higher proportion of students may not reach a standard score of 90 on fluency and comprehension measures than on word reading. This finding is consistent with prior RTI research (Denton et al., 2006; Denton et al., 2013; O'Connor et al., 2005; Vellutino et al., 2006; Authors, 2008); an implication of this finding is the need to develop interventions with greater potency to achieve and to sustain catch up growth in the area of fluency and comprehension.

We also found the proportion of students with standard scores below a 90 one and two years following intervention was much lower in the present study than in previous research (in which proportions of students ranged from 28% to 60% for students who received Tier 2 and Tier 3). It is important to consider this finding in light of difference across studies that could reflect different selection criteria. For example, not all SR students received Tier 3 and therefore may represent a less impaired sample than in other studies. However, it was encouraging that even among SR students, by the end of second and third grade only 8.7% and 7.9%, respectively, scored below a 90 in word reading. Furthermore, by the end of third grade, none of the students who were less difficult to remediate in first grade demonstrated word reading scores below a 90.

In addition, we found relatively larger proportions of dysfluent readers or poor comprehenders than inaccurate word readers (by third grade 42.1%, 29.4% and 24.5% of students who were difficult to remediate scored below a 90 on word reading efficiency, oral reading fluency, and passage comprehension, respectively). This pattern is similar to findings in a prior Tier 2 research study (e.g., Fletcher et al., 2011) and in a recent Tier 3 study (Denton et al., 2013), which reported that, on average, students who received Tier 3 had oral reading fluency scores below the 20^{th} percentile for their grade level and 64% did not achieve a standard score of 90 on passage comprehension. This finding is also consistent with Vellutino et al.'s (2006) hypothesis that inadequate responders demonstrate a pattern along a continuum of severity on a variety of reading-related variables, particularly for students who might be assumed to represent the most persistently inadequate responders.

Fourth, it is noteworthy that unlike two previous longitudinal follow up studies (Gilbert et al., 2013; Vellutino, 2006), we found that with the exception of word reading efficiency, there were no important increases in the proportion of students with reading skills below a 90 from second to third grade. Specifically, by third grade Gilbert et al reported 46% of students who received Tier 3 in first grade had word reading and comprehension skills

below a 90 and Vellutino et al found roughly a third of students who were difficult to remediate had basic reading skills below a 90. An implication may be that interventions should be more tailored for SR students and should include more fluency and comprehension training.

Limitations and Directions for Future Research

There are a number of important limitations of this study. First, as is common in any longitudinal school based research, we experienced attrition. However, as we noted previously, there were no differences between the attrited and non attrited groups on race, free and reduced lunch or entry literacy skills. Second, although intervention groups were nested within interventionists, these groups changed each 8 weeks per the design of the study and so, due to statistical complexity, this level of variance was not accounted for in the analysis. Third, although we carefully observed Tier 1 and documented the fidelity of Tier 2 and 3 implementation in first grade, we do not have information about the instruction or intervention received by students in second or third grade. While second and third grade teachers used the same core instructional program for Tier 1, we did not observe this, nor do we have detailed information about interventions or who received them beyond the time frame of our study. However, we will explore in future research whether schools have identified students for special education for learning disabilities and that examines their cognitive and academic profiles associated with their responsiveness. Fourth, we did not examine a no-treatment control group so it is not possible to identify students who may have succeeded without intervention. In the framework of this study and the work of the schools it would have been unethical to deny intervention to a representative group of struggling readers. Finally, the generalizability of our findings may be limited to strong implementation in Tier 1, aligned standard protocol interventions implemented with fidelity by project staff, and to similar populations. In addition, findings may have differed if we had used other interventions or other criteria for adequate response and moving up and down tiers. Future research is needed to replicate the Dynamic model using various interventions, including a broader range of schools, and involving teachers and special educators as interventionists.

Acknowledgments

This work was supported by grant P50HD052120 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD or the National Institutes of Health.

References

Beach KD, O'Connor RE. Early response to intervention measures and criteria as predictor of reading disability in the beginning of third grade. Journal of Learning Disabilities. 2013 Doi: 10.1177/0022219413495451.

Benjamini Y, Hochberg Y. Journal of the Royal Statistical Society B. 1995; Controlling the false discovery rate - a new and powerful approach to multiple testing.57:289–300.

Bereiter, C.; Brown, A.; Campione, J.; Carruthers, I.; Case, R.; Hirshberg, J.; Treadway, GH. Open court reading. SRA McGraw-Hill; Columbus, OH: 2002.

Bradley, R.; Danielson, L.; Hallahan, R.; R., editors. Identification of learning disabilities: Research to practice. Erlbaum; Mahwah, NJ: 2002.

Denton CA, Fletcher JM, Anthony JL, Frances DJ. An Evaluation of Intensive Intervention for Students with Persistent Reading Difficulties. Journal of Learning Disabilities. 2006; 39(5):447–466. doi: http://dx.doi.org/10.1177/00222194060390050601. [PubMed: 17004676]

Denton CA, Tolar TD, Fletcher JM, Barth AE, Vaughn S, Francis DJ. Effects of tier 3 intervention for students with persistent reading difficulties and characteristics of inadequate responders. Journal of Educational Psychology. Aug. 2013; 1052013(3):633–648. doi: 10.1037/a0032581. [PubMed: 25308995]

Fletcher JM, Coulter WA, Reschly DJ, Vaughn S. Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. Annals of Dyslexia. 2004; 54:304–331. [PubMed: 15741940]

Fletcher JM, Stuebing KK, Barth AE, Denton CA, Cirino PT, Francis DJ, Vaughn S. Cognitive correlates of inadequate response to reading intervention. School Psychology Review. 2011; 40(1): 3–22. [PubMed: 23125475]

Foorman BR, Francis DJ, Fletcher JM, Schatschneider C, Mehta P. The role of instruction in learning to read: Preventing reading disabilities in at-risk children. Journal of Educational Psychology. 1998; 90:37–55. doi: http://dx.doi.org/10.1037//0022-0663.90.1.37.

Fuchs LS, Fuchs D, Compton DL. Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. Exceptional Children. 2004; 71(1):7–21.

Gersten, R.; Compton, D.; Connor, CM.; Dimino, J.; Santoro, L.; Linan-Thompson, S.; Tilly, WD. Assisting Students Struggling with Reading: Response to Intervention (RtI) and Multi-Tier Intervention in the Primary Grades. 2009. Retrieved from http://ies.ed.gov/ncee/wwc/PracticeGuide.aspx?sid=3

Gilbert JK, Compton DL, Fuchs D, Fuchs LS, Bouton B, Barquero LA, Cho E. Efficacy of a First-Grade Responsiveness-to-Intervention Prevention Model for Struggling Readers. Reading Research Quarterly. 2013; 48(2):135–154. doi: http://dx.doi.org/10.1002/rrq.45.

Good, RH.; Kaminski, RA. Dynamic indicators of basic literacy skills. 6th ed.. University of Oregon, Institute for the Development of Educational Achievement; Eugene, OR: 2001.

Haager, D.; Gersten, R.; Baker, S.; Graves, A. The English language learner observation instrument for beginning readers.. In: S. V.; Briggs, KL., editors. Reading in the classroom: Systems for the observation of teaching and learning. Brookes; Baltimore, MD: 2003.

Hasbrouck J, Tindal GA. Oral reading fluency norms: A valuable assessment tool for reading teachers. The Reading Teacher. 2006; 59(7):636–644.

Hill DR, King SA, Lemons CJ, Partanen JN. Fidelity of implementation and instructional alignment in Response to Intervention research. Learning Disabilities Research & Practice. 2012; 27(3):116–124. doi:10.1111/j.1540-5826.2012.00357.x.

Lam EA, McMaster K. Predictors of responsiveness to early literacy intervention: A ten year update. Manuscript accepted for publication in Learning Disabilities Quarterly. (in press).

Little RJA. A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association. 1988; 83:1198–1202.

Lonigan CJ, Whitehurst GJ. Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. Early Childhood Research Quarterly. 1998; 17:265–292. doi: http://dx.doi.org/10.1016/S0885-2006(99)80038-6.

Lyon GR, Chhabra V. The current state of science and the future of specific reading disability. Ment. Retard. Dev. Disabil. Res. Rev. 1996; 2:2–9. doi: 10.1002/(SICI)1098-2779(1996)2:1<2::AID-MRDD2>3.0.CO;2-X.

Mathes PG, Denton CA, Fletcher JM, Anthony JL, Francis DJ, Schatschneider C. The effects of theoretically different instruction and student characteristics on the skills of struggling readers. Reading Research Quarterly. 2005; 40(2):148–182. doi:10.1598/RRQ.40.2.2.

Mathes PG, Torgesen JK, Wahl M, Menchetti JC, Grek ML. Proactive beginning reading: Intensive small group instruction for struggling readers. Developed with funds provided by the National Institute of Child Health and Human Development (#R01 HD). Prevention and Remediation of Reading Disabilities. 1999

National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. 2000. Available on-line: http://www.nichd.nih.gov/publications/nrp/smallbook.htm

Nelson, J.; Benner, GJ.; Gonzalez, J. Learning Disabilities Research & Practice. Vol. 18. Blackwell Publishing Limited; 2003. Learner Characteristics that Influence the Treatment Effectiveness of Early Literacy Interventions: A Meta-Analytic Review.; p. 255doi:10.1111/1540-5826.00080

O'Connor R. Increasing the intensity of intervention in kindergarten and first grade. Learning Disabilities Research and Practice. 2000; 15(1):43–54. Retrieved from http://search.proquest.com/docview/62443452?accountid=4840.

O'Connor RE, Harty KR, Fulmer D. Tiers of intervention in kindergarten through third grade. [RtI]. Journal of Learning Disabilities. 2005; 38(6):532–538. doi: 10.1177/00222194050380060901. [PubMed: 16392695]

Shinn, MM.; Shinn, MR. AIMSweb. Edformation; Eden Prairie, MN: 2004.

Simmons, DC.; Kameenui, EJ.; Stoolmiller, M.; Coyne, MD.; Harn, B. Accelerating growth and maintaining proficiency: A two-year intervention study of kindergarten and first-grade children at risk for reading difficulties.. In: Foorman, B., editor. Preventing and remediating reading difficulties: Bringing science up to scale. York Press; Timonium, MD: 2003. p. 197-228.

Speece DL, Schatschneider C, Silverman R, Case LP, Cooper DH, Jacobs DM. Identification of reading problems in first grade within a response-to-intervention framework. The Elementary School Journal. 2011; 11:585–607. doi: http://dx.doi.org/10.1086/659032. [PubMed: 22058431]

Torgesen, JK.; Wagner, RK.; Rashotte, CA. Test of word reading efficiency. PRO-ED; Austin, TX: 1997.

Torgesen, JK.; Wagner, RK.; Rashotte, CA. Test of Word Reading Efficiency. 2nd Edition.. Pro-ED, Inc.; Austin, TX: 2012.

Torgesen JK, Wagner RK, Rashotte CA, Rose E, Lindamood P, Conway T, Garvan C. Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. Journal of Educational Psychology. 1999; 91:579–593.

Tran L, Sanchez T, Arellano B, Swanson LH. A meta-analysis of the RTI literature for children at risk for reading disabilities. Journal of Learning Disabilities. 2011; 44(3):283–295. Retrieved from http://search.proquest.com/docview/870288217?accountid=4840. [PubMed: 21521870]

Vaughn S, Denton CA, Fletcher JM. Why intensive interventions are necessary for students with severe reading difficulties. Psychology in the Schools. 2010; 47(5):432–444. [PubMed: 21072127]

Vaughn, S.; Wanzek, J.; Linan-Thompson, S.; Murray, CS. Monitoring response to supplemental services for students at risk for reading difficulties: High and low responders.. In: Jimerson, SR.; Burns, MK.; VanDerHeyden, AM., editors. Handbook of response to intervention: The science and practice of assessment and intervention. Springer Science + Business Media; New York, NY US: 2007. p. 234-243.doi:10.1007/978-0-387-49053-3_17

Vaughn S, Wanzek J, Murray CS, Scammacca N, Linan-Thompson S, Woodruff AL. Response to early reading intervention: Examining higher and lower responders. Exceptional Children. 2009; 75(2):165–183.

Vellutino F, Scanlon D, Sipay E, Small S, Pratt A, Chen R, et al. Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. Journal of Educational Psychology. 1996; 88:601–638.

Vellutino FR, Scanlon DM, Small SG, Fanuele DP. Response to Intervention as a Vehicle for Distinguishing Between Reading Disabled and Non-Reading Disabled Children: Evidence for the Role of Kindergarten and First Grade Intervention. Journal of Learning Disabilities. 2006; 39:2, 157–169. doi: http://dx.doi.org/10.1177/00222194060390020401. [PubMed: 16512078]

Vellutino FR, Scanlon DM, Zhang H, Schatschneider C. Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. Reading and Writing. 2008; 21(4):437–480. doi:http://dx.doi.org/10.1007/s11145-007-9098-2.

Wechsler, D. Wechsler Individual Achievement Test. Third Edition.. NCS Pearson; San Antonio, TX: 2009.

Woodcock, R.; McGrew, K.; Mather, N. Woodcock Johnson-III tests of achievement. Riverside; Itasca, IL: 2001.

J Res Educ Eff. Author manuscript; available in PMC 2014 October 23.

Table 1

	No Risk			Easy to Remediate			Sustained Remediattion		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Fall Grade 1									
LWID raw	261	28.70	8.22	31	18.55	3.62	126	19.49	3.89
LWID SS	262	109.06	14.00	31	92.84	9.86	126	92.67	10.01
TOWRE SWE raw	262	26.23	16.06	31	7.29	4.62	126	8.52	5.97
TOWRE SWE SS	262	102.27	13.27	31	86.52	9.98	125	86.44	9.94
TOWRE PDE raw	262	9.69	8.55	31	1.87	3.00	126	1.92	2.90
TOWRE PDE SS	262	102.05	12.08	31	90.48	10.50	126	90.02	8.53
TOWRE combined SS	262	102.59	14.61	31	86.23	11.77	125	85.91	10.30
ORF median	262	40.85	30.60	30	11.47	6.14	126	10.38	6.53
Passage comp raw	261	13.28	5.18	31	6.45	1.93	126	7.59	2.71
Passage comp SS	262	100.20	14.83	31	82.48	8.16	126	83.47	13.84
Spring Grade 1									
LWID raw	259	40.81	6.43	31	35.10	4.78	126	30.62	5.46
LWID SS	259	115.04	10.75	31	107.94	10.82	126	99.17	11.05
TOWRE SWE raw	259	50.79	12.66	31	38.68	13.25	126	27.98	11.01
TOWRE SWE SS	259	103.87	13.83	31	93.55	15.51	126	83.72	12.56
TOWRE PDE raw	259	21.42	10.95	31	15.35	8.53	126	8.05	5.58
TOWRE PDE SS	259	100.92	13.14	31	94.68	11.89	126	84.53	9.64
TOWRE combined SS	259	102.39	12.90	31	94.11	13.06	126	84.13	10.39
ORF median	259	86.03	36.93	31	54.19	30.32	126	30.94	18.63
Passage comp raw	259	21.01	4.57	31	17.06	3.60	126	14.63	3.67
Passage comp SS	259	104.15	11.01	31	96.52	11.03	126	89.50	10.89
Spring Grade 2									
LWID raw	225	47.64	7.14	23	43.35	6.49	104	39.32	5.03
LWID SS	225	110.91	10.90	23	106.39	11.63	104	98.59	8.97
TOWRE SWE raw	222	61.32	9.92	23	54.96	11.00	104	47.13	11.77
TOWRE SWE SS	222	111.86	11.68	23	105.65	12.45	104	96.50	12.62
TOWRE PDE raw	222	30.53	10.41	23	25.04	10.31	104	17.38	8.84

	No Risk			Easy to Remediate			Sustained Remediattion		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
TOWRE PDE SS	222	109.41	12.76	23	103.70	13.93	104	93.73	11.01
TOWRE combined SS	222	112.77	13.71	23	105.57	14.97	104	94.12	13.13
ORF median	223	108.62	33.60	23	90.80	30.35	103	68.63	25.03
Passage comp raw	225	24.88	4.19	23	22.65	3.27	104	20.79	3.97
Passage comp SS	225	100.00	9.76	23	96.52	7.90	104	90.79	9.92
Spring Grade 3									
LWID raw	161	53.04	5.87	15	50.13	4.02	85	45.29	5.36
LWID SS	161	109.30	9.76	15	106.00	7.50	85	96.92	8.99
TOWRE SWE raw	165	67.25	8.77	18	68.50	6.76	85	53.49	11.45
TOWRE SWE SS	165	101.72	12.16	18	104.22	9.75	85	84.38	13.43
TOWRE PDE raw	165	35.93	10.47	18	36.44	10.04	85	22.16	10.48
TOWRE PDE SS	165	103.36	13.61	18	104.39	13.18	85	85.35	14.19
TOWRE combined SS	165	102.30	12.83	18	104.56	11.67	85	84.12	13.62
WIAT ORF Weighted raw	168	116.63	33.73	18	114.94	31.25	84	80.79	26.52
WIAT ORF SS	168	108.42	13.63	18	108.00	12.62	84	93.00	11.66
Passage comp raw	161	26.72	3.25	15	25.20	3.00	85	24.05	3.29
Passage comp SS	161	97.96	8.71	15	94.93	7.14	85	90.74	9.10

Note: LWID = Woodcock Johnson-III Letter Word Identification; TOWRE = Test of Word Reading Efficiency; ORF = Oral reading fluency; WIAT = Wechsler Individual Achievement Test; Passage comp = Woodcock Johnson-III Passage Comprehension

Table 2

Results of multi-level model for reading measures in second grade predicted by intervention condition, group status, and interaction between intervention condition and group status

	Letter Word Identification	TOWRE	Oral reading fluency	Passage Comp
Fixed Effects				
Intercept	49.02 (3.55)***	46.78 (3.71)***	87.25 (4.69)***	60.67 (3.30)***
Male	2.40 (.76)**	2.58 (1.18)*	−3.77 (2.53)	1.39 (.81)
Free and reduced lunch	−.25 (.98)	−.55 (1.30)	−5.59 (2.67)*	−2.28 (.86)**
Black	−1.65 (.98)	.63 (1.02)	−3.69 (2.40)	−2.64 (.85)**
Others	−.24 (1.84)	1.42 (2.29)	.54 (6.33)	−2.88 (1.48)*
Grade 1 fall score	.57 (.02)***	.63 (.03)***	.70 (.04)***	.42 (.03)***
Dynamic condition	−.39 (1.01)	−.04 (1.28)	−.97 (3.29)	−1.06 (.94)
Easy to remediate (ER)	−7.31 (4.47)	−18.13 (2.61)***	−40.49 (6.82)***	−4.51 (4.24)
Sustained remediation (SR)	−3.92 (.97)***	−9.63 (1.76)***	−17.07 (4.92)***	−2.21 (1.19)
Dynamic*ER	14.00 (5.15)**	23.86 (3.13)***	51.64 (8.59)***	10.62 (4.54)*
Dynamic* SR	.94(1.78)	1.65 (2.78)	−1.73 (6.40)	.14 (1.87)
Variance Components				
Classroom	1.33	2.99	16.46	0
Individuals	46.89***	102.43***	550.24***	49.49***

Note: TOWRE = Test of Word Reading Efficiency; Passage comp = Passage comprehension

Table 3

Results of multi-level model for reading measures in third grade predicted by intervention condition, group status, and interaction between intervention condition and group status

	Letter Word Identification	TOWRE	Oral reading fluency	Passage Comp
Fixed Effects				
Intercept	60.11 (4.21)***	50.62 (4.90)***	101.39 (2.68)***	67.36 (2.93)***
Male	1.30 (.85)	.29 (1.37)	−2.41 (1.38)	1.27 (.82)
Free and reduced lunch	−1.08 (1.02)	−1.30 (1.48)	−1.88 (1.51)	−1.18 (.87)
Black	−1.52 (1.11)	.93 (1.47)	−1.09 (1.61)	−3.83 (1.06)***
Others	−1.18 (1.57)	3.02 (2.12)	.51 (2.59)	−2.67 (1.50)
Grade 1 fall score	.45 (.04)***	.51 (.04)***	.23 (.03)***	.33 (.03)***
Dynamic condition	−.24 (1.10)	−1.63 (1.68)	−.56 (1.75)	1.03 (1.00)
Easy to remediate (ER)	5.43 (1.97)**	10.11 (2.90)***	7.12(3.20)*	2.83 (1.75)
Sustained remediation (SR)	−4.19 (1.42)**	−10.54 (2.03)***	−6.17(2.23)**	−.03 (1.33)
Dynamic* SR	.08 (1.80)	1.75 (2.68)	−2.94 (3.18)	−2.77 (1.86)
Variance Components				
Classroom	1.26	2.38	14.63	2.97
Individuals	48.59***	115.46***	105.38***	41.31***

Note: TOWRE = Test of Word Reading Efficiency; Passage comp = Passage comprehension

Table 4

Percentage of students whose standard scores are below 90 in reading measures

	No risk n = 261	Easy to remediate n = 31	Sustained remediation n = 126
End of Grade 2			
WJ Letter Word Identification	4 (1.5%)	1 (3.2%)	11 (8.7%)
Test of Word Reading Efficiency	9 (3.4%)	3 (9.7%)	36 (28.6%)
Oral reading fluency	11 (4.2%)	4 (12.9%)	36 (28.6%)
WJ Passage comprehension	31 (11.8%)	4 (12.9%)	46 (36.5%)
All the reading measures	1 (.4%)	1 (3.2%)	9 (7.1%)
End of Grade 3			
WJ Letter Word Identification	3 (1.1%)	0 (0%)	10 (7.9%)
Test of Word Reading Efficiency	25 (9.5%)	1 (3.2%)	53 (42.1%)
Oral reading fluency	14 (5.3%)	1 (3.2%)	37 (29.4%)
WJ Passage comprehension	22 (8.4%)	4 (12.9%)	31 (24.5%)
All the reading measures	2 .8%)	0 (0%)	6 (4.8%)