Published in final edited form as:

Learn Disabil Q. 2014 November ; 37(4): 204–217. doi:10.1177/0731948714526999.

Understanding Inadequate Response to First Grade Multi-Tier Intervention: Nomothetic and Idiographic Perspectives

Luana Greulich, Stephanie Al Otaiba, Christopher Schatschneider, Jeanne Wanzek, Miriam Ortiz, and Richard Wagner

Abstract

The purpose of this study was to use a mixed methods approach to learn about inadequate response to a year-long multi-tier RTI model that allowed first-grade students to move up and down tiers. Participants were 156 students who received supplemental intervention services during a larger multi-tier RTI study involving classrooms and 522 students across 10 schools. Findings from an all-subset regression indicate letter word reading, the fluency composite, and blending words explained the most variance (15%) in response among initial skills. Adding additional teacher ratings of behavior and academics, accounted for a small amount of additional variance (3%) in group membership. The ROC curve analysis indicated 87.5% of students were correctly classified, yielding a sensitivity of 85.3 and a specificity of 65.0. Findings from qualitative observations of intervention sessions suggest inadequate responders demonstrated physical and verbal task avoidance and displayed emotions of hopelessness and shame. Implications for practice are discussed.

Understanding responsiveness to intensive interventions, including Tier 3, within multi-tier models of Response to Intervention, RTI, is important for the learning disability community because many states are still in the process of developing their RTI models and because Tier 3 varies across, and even within states (e.g., Berkeley, Bender, Peaster & Saunders, 2009; Fuchs, Fuchs & Compton, 2012; Fuchs, Fuchs & Stecker, 2010; Fuchs & Vaughn, 2012; Mellard, McKnight & Jordan, 2010; Mellard, McKnight & Woods, 2009). A review of RTI implementation across states emphasized that only one state (Delaware) had explicit criteria for when students should move in and out of Tier 3; only six states had explicit guidelines for referral to special education for students in the RTI process (Zirkel & Thomas, 2010). This is concerning because RTI is encouraged, if not mandated, not only for prevention purposes, but also for identification of students with learning disabilities. Consequently, RTI is an important foundation of a free and appropriate public education (FAPE) guaranteed to all students under federal legislation like IDEA 2004.

Furthermore, as Lam and McMaster (2013) report in their literature review in the present issue, although from a prevention perspective, RTI models may reduce the need for remedial services, converging evidence indicates some students will need special education. A number of research syntheses examining RTI describe students who do not show adequate response to generally effective Tier 1 and Tier 2 interventions (Al Otaiba & Fuchs, 2002; Burns, Appleton & Stehouwer, 2005; Fuchs, Mock, Morgan & Young, 2003; Gersten et al., 2009; Hill, King, Lemons, & Partanen, 2012; Nelson, Benner, & Gonzalez, 2003; Tran, Sanchez, Arallano & Swanson, 2011; Wanzek & Vaughn, 2007). However, as noted by Lam

and McMaster (2013), in comparison with solid evidence about Tier 2 interventions, there have been relatively fewer studies examining response to multi-tier models that include Tier 3.

To our knowledge, there have been eight experimental or quasi-experimental studies involving Tier 3 (Al Otaiba et al., in press; Denton, Fletcher, Anthony & Francis, 2006; Denton et al., 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).

The first study to include Tier 3 and to incorporate fluid movement up or down all three tiers based upon student data was conducted by O'Connor and colleagues (O'Connor, Harty, & Fulmer, 2005), who used a complex longitudinal quasi-experimental design (kindergarten through third grade). It was promising that most students who received Tier 2 intervention achieved grade level standards, only 10 needed Tier 3 (out of the initial 31 students), but it was discouraging that of these, only four could read on grade level by the end of third grade; thus 60% were inadequate responders to Tier 3.

Several additional research teams provided Tier 3 to students and included students who did not respond to a year-long effective first grade Tier 1 and Tier 2. None, however, allowed movement up or down tiers during the study year. Denton et al. (2006) provided intensive Tier 3 intervention to 27 students (half of whom had participated in a prior study and received Tier 1 or Tier 2 but half had not). More than half (15) demonstrated little or no growth, defined by the study authors as gaining less than 0.5 standard score points on a composite score of basic reading skills (word identification and word attack); nine were third graders and seven already had special education labels. In another more recent study, Denton and colleagues (Denton et al., 2013) randomly assigned students to research intervention vs. school-delivered intervention and reported significantly greater growth for students in the research intervention condition on word identification, decoding, and word reading fluency (with no differences on oral reading fluency or comprehension or on the percent of students who achieved grade level benchmarks). Denton and colleagues reported highly variable individual differences in response and that a higher percentage of students performed below a standard score of 90 (43%) on word reading fluency and comprehension (36%) than on word reading (72%). Inadequate responders had weaker language and overall reading skills, and the authors suggested the need for further study of the characteristics of students who show inadequate response to Tier 3.

Vaughn and colleagues (Vaughn et al., 2007; Vaughn et al., 2009) provided Tier 3 only to students who had not responded to Tier 1 and 2 the prior year. Although their Tier 3 intervention addressed all areas of reading, statistically significant differences in gains compared to controls were only seen on untimed word reading and reading comprehension. A significant interaction with oral reading fluency (ORF) indicated that students with higher initial ORF scores (above 40 words per min at the start of second grade) grew more than students with lower ORF scores.

Additional concerns about RTI from this research base stem from evidence that standard scores decline following kindergarten or first grade Tier 2 and Tier 3 interventions for some students. Vellutino et al.'s (2006) findings indicated that the students who were most difficult to remediate could not maintain gains after receiving Tiers 2 and 3; by third grade, nearly a third read below the 30th percentile. Gilbert et al. (2013) found even higher proportions of students who were nonresponders to Tier 2 reading below this benchmark at third grade after receiving either additional Tier 2 or a Tier 3 (60% and 46%, respectively). Further, although the nonresponders in both tiers received the same type of intervention, Tier 3 was delivered one-to-one, yet no significant differences favored Tier 3.

The final study was recently completed by our own research team comparing two models of RTI in reading for first grade students (Al Otaiba, in press). Details are described in the methods section, but briefly, one study group fast-tracked students to intervention and the other used a two-stage process that required students to begin in Tier 1 and only receive intervention if they did not respond. Tier 1 was noted as generally effective and Tiers 2 and 3 were implemented with fidelity. It is relevant to know that a small percentage of our students ended first grade with word reading standard scores below 91 (4.30% on word attack and 7.20% on letter word identification); however, the percentage was 19.40% on passage comprehension.

All of these studies incorporated several of the five core components deemed essential for RTI according to Gersten et al.'s (2009) *Practice Guide for Response to Intervention* (i.e., universal screening, a high quality core reading program, progress monitoring, increasingly intensive tiers of intervention, and fidelity of implementation). Those components with less research are noteworthy, in particular the effectiveness of Tier 1 has been documented rarely (Hill et al., 2012). Moreover, despite calls for layers of increasing intensive intervention, in only three studies have students been allowed to move across tiers within a study year and only two allowed movement up or down a tier. Generally, findings from these studies provide promise for Tier 3 for some students not helped by Tiers 1 and 2, but yield concern that a substantial proportion of persistently inadequate responders will not catch up to grade level reading performance on at least some measures of reading (e.g., comprehension, fluency, phonemic decoding) and will likely need ongoing support. Further, there are concerns that students with the weakest initial skills may remain in Tier 2 for too long, thus making RTI another potential wait to fail system. Finally, the findings indicate the need to better understand individual differences in response to intervention across various measures and to explore the characteristics of students who do not respond to multi-tier interventions.

The purpose of the present study stems from prior research on responsiveness and borrows from an article by Fuchs et al. (2003), who argued the value of considering both a nomothetic and an idiographic perspective (Cone, 1986). A nomothetic perspective examines, on average, which variables predict classes of behavior, in the present study adequate and inadequate response. An idiographic perspective involves a qualitative examination or in-depth description of individuals; in the present study we wanted to understand engagement and emotions observable during intervention sessions that might distinguish inadequate and adequate responders, due in part, to findings from a mixed methods longitudinal study of RTI conducted by Case, Speece and Malloy (2003). Case et

al. found that some inadequate responders became responsive to intervention when they had a teacher who motivated them to stay engaged or who encouraged their sense of pride. Researchers have linked motivation and engagement to the process of becoming a proficient reader (e.g., Gambrell, Palmer, Codling & Mazzoni, 1996; Onatsu-Arvilommi & Nurmi, 2000; Wigfield & Guthrie, 1997). Onatsu-Arvilommi et al. specifically noted a negative relation between task avoidance and reading performance in first grade. Quirk and Schwanenflugel (2004) noted that remedial reading interventions should consider motivation. Morgan and Fuchs (2007) reviewed the correlational research regarding reading motivation and skill and reported that several studies suggested a bidirectional relationship. Motivation has been defined in a variety of ways, and we chose to focus specifically on engagement/avoidance. Pekrun and colleagues (Pekrun, Elliot, & Maier, 2006; Pekrun, Elliot, & Maier, 2009) have proposed that what they term *achievement emotions* (enjoyment, boredom, anger, hope, pride, anxiety, hopelessness and shame) may have an important relation with academic achievement among older students. They suggest that positive emotions such as pride, hope and enjoyment lead to better motivation, self-regulation and higher task engagement. By contrast, negative emotions such as anger, anxiety, boredom, hopelessness, and shame may reduce cognitive resources and self-regulation and would likely lead to task avoidance. Thus, it is important to examine these emotions from an individual differences perspective.

Purpose and Research Questions

The present study took a mixed methods approach to learn from a nomothetic perspective more about whether differences in initial skills and ratings of behavior predict adequate and inadequate response to a year-long multi-tier RTI model that allows students to move up and down tiers. Our first research question asked: Can initial language and literacy skills, and teacher judgment about behavior predict adequate and inadequate response to intervention? The study also used an idiographic qualitative approach to address engagement/avoidance behavior and emotions displayed by inadequate responders during intervention to guide thinking about possible malleable targets for improving and individualizing interventions. Thus, the second question asked: Were there observable differences between inadequate and adequate responders in terms of their engagement/avoidance and academic emotions during intervention sessions?

METHOD

Background of Larger Study

The present study focused on the 156 students who received supplemental intervention (Tier 2 or Tier 3) services as part of a larger just-mentioned multi-tier RTI study (Al Otaiba et al., in press). That study involved 34 first-grade classrooms and 522 students across 10 schools in a southeastern school district and was part of the Florida Learning Disabilities Center Grant. Across these schools there was a strong focus on first grade Tier 1 reading; core reading instruction (*Open Court*; Bereiter et al., 2002) was provided for a minimum of 90 min. In one school only 15.8% of students were part of the free and reduced lunch (FARL) programs, but in the remaining schools, FARL status ranged from 42.8% to 89.9%; small percentages of students had Limited English Proficiency (ranging from 0.4% to 2.8%).

On average, the 34 participating classroom teachers were experienced ($M = 14.54$ years teaching; $SD = 9.74$), nearly a third held graduate degrees ($n = 9$), and all agreed that their language arts block be videotaped twice during the year to observe the quality of Tier 1 instruction, which was rated as effective. Screening for students' eligibility for supplemental intervention in the larger study began in September of first grade using four screeners of letter sound fluency, word reading fluency, and word attack fluency (see measures). The 40th percentile was selected to represent grade level performance. We also asked teachers to rate the severity of students' reading difficulties relative to classmates (Speece et al., 2011). Following the initial screening, students were considered eligible for Tier 3 if their teachers reported they had severe reading difficulties and they scored below grade level (the 40th percentile at the school level) on all four screeners. Students were considered eligible for Tier 2 if their teachers reported they had severe reading difficulties or they scored below the grade level (below the 40th percentile) at the school level for three out of four screeners. Students with word identification and passage comprehension scores above a 95 were excluded from supplemental intervention.

Next, we calculated z -scores on the screeners, averaged them, rank ordered students within tier eligibility status within classroom, identified adjacent pairs in the ranking, and randomly assigned one member of the pair to one of two RTI conditions: (a) Typical RTI which followed two-stage RTI decision rules (start in Tier 1 and sequentially move to Tier 2 or 3 as needed); or (b) Dynamic RTI, which fast-tracked students to Tier 2 or Tier 3 interventions immediately according to initial screening results. Then, after 8 weeks, children were again screened and 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. When students were successful (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, then they received a less intensive tier.

All supplemental interventions were identical across conditions and were provided by well-trained interventionists (advanced graduate students and certified teachers) who also were trained in positive behavior management. For Tier 2, students received two, 30 min weekly sessions in groups of 4–7. For Tier 3, students received four, 45 min weekly sessions in groups of 1–3. For Tier 2, code-focused activities were drawn from the first grade Open Court *Imagine It!* (Berierter et al., 2002) series and the Florida Center for Reading Research K-3 Center Activities (www.fcr.org) and included phonological awareness and letter sound skills, decoding and sight word instruction, and fluency training. Across both Tier 2 and Tier 3, interventionists also provided meaning-focused instruction for about 10–15 min per day. In the first 8 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 Tier 3, code-focused activities followed the standard Early Interventions in Reading (EIR; Mathes, Torgesen, Wahl, Menchetti, & Grek, 1999) that includes: phonemic awareness,

alphabetics and phonics, and fluency (roughly 30 min). For fidelity purposes, interventionists videotaped intervention sessions every 8 weeks.

Participants in the Present Study

For the present study, we focused on the 156 students who received Tier 2 or 3 intervention during their first grade year. We applied a cut-point for inadequate response of performance at or below the 25th percentile of intervention students on the WJ-III Basic Reading Skills Cluster (Woodcock, McGrew, & Mather, 2001) at the end of the year after intervention (using a similar end of study cut-point as some other researchers have used Fletcher et al., 2011). A total of 20 of the 156 students were thus considered inadequate responders (Table 1 shows the Tiers received by response status). There were 10 inadequate responders in each condition. A total of 44 adequate responders were in the Typical condition and 92 were in the Dynamic condition. Chi square analyses indicated no significant differences between adequate and inadequate responders respectively on FARL (48.53% and 75%; $\chi^2 (2, N = 156) = .314, p = .13$), Race (African American 59.55% and 70%, Caucasian 33.08 and 15%, Multi-racial 2.95% and 5%, Hispanic 1.47% and 10%, and Asian 2.94% and 0%; $\chi^2 (4, N = 156) = 8.09, p = .09$, on limited English proficiency (1.47% and 0% $\chi^2 (1, N = 156) = .29, p = .59$), or gender (62.5% males and 65% males $\chi^2 (1, N = 156) = .047, p = .83$). An ANOVA revealed no differences in attendance between adequate or inadequate responders $F(1,168) = .426, p = .515 (M = 8.69 \text{ and } 9.91)$.

Measures and Data Collection Procedures

All data used in the present study had been collected in the larger study by a highly trained research team. The 16 assessments we used to discern differences between adequate and inadequate responders are categorized into constructs of phonological awareness, vocabulary, untimed reading, fluency, reading, IQ, and teacher judgment about reading, behavior, and attention. Screeners are described within the fluency composite and noted with an *.

Phonological awareness—To measure phonological awareness, the subtests of Blending and Elision of the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) were given in late fall. The Blending subtest of the CTOPP requires the students to listen to a series of real words spoken in segments and earn a point for each word they blend correctly. The Elision subtest of the CTOPP requires a student to listen to a segmented word and then pronounce the word without a segment. Internal consistency for each subtest exceeds .80; test-retest coefficients range from .70 to .92. Standard scores were used.

Expressive vocabulary—The subtest of Picture Vocabulary of the Woodcock-Johnson III Test of Achievement (WJ-III; Woodcock et al., 2001) was used to measure the construct of language and was given three times during the year (fall, winter, and spring). The Picture Vocabulary subtest measures student's expressive vocabulary by asking them to name pictured objects. Median reliability is .77. W-scores were used for analytical purposes.

Untimed reading—The construct of reading was assessed using the subtests Letter Word Identification, Word Attack and Passage Comprehension of the Woodcock-Johnson III Test of Achievement (WJ-III; Woodcock et al., 2001); reliabilities are .91, .87, and .83, respectively. The Letter-Word Identification subtest consists of students fluently reading words that increase in difficulty. The subtest Word Attack, measures the ability to decode nonsense words of increasing difficulty. The subtest Passage Comprehension measures comprehension by asking students to identify a missing word in a written passage; this is known as a cloze procedure. W-scores from were used for analytical purposes.

Fluency Composite—Fall scores were used from several fluency measures to create a composite (the first three assessments were screeners). The AIMSweb (Shinn & Shinn, 2004) Letter Sound Fluency* requires students to quickly and accurately name the given letter sounds within 1 minute. Alternate form reliability is .90. Raw scores were used for analytical purposes. The two subtests Sight Word Efficiency* and Phonemic Decoding Efficiency* of the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1997b) require the student to read a list of sight words and a list of pseudo-words that increase in difficulty in 45 seconds. Alternate forms and test-retest reliability coefficients for each subtest exceed .90 for this age of student. Standard scores were used. The Word Identification Fluency* (WIF; Fuchs, Fuchs, & Compton, 2004) requires students to read a list of high frequency words randomly ordered for 1-minute; alternate form reliability is .97. Raw scores were used for analytical purposes. The Oral Reading Fluency subtest of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2001) is a brief, 1-minute measure where students are asked to read a grade-level passage and the number of correct words per minute is recorded; three passages are given for the assessment and for this study, mean raw scores were analyzed. Test-retest reliability for elementary students is .92.

IQ—The Kaufman Brief Intelligence Test 2 (KBIT; Kaufman & Kaufman, 2004) was used to determine an informal IQ score. The KBIT was given one time during the year either in the fall or winter depending upon when the student entered intervention. The KBIT is individually administered and measures both verbal and nonverbal intelligence that includes three subtests: Verbal Knowledge and Riddles (Verbal intelligence), and Matrices (Non-verbal intelligence). The average reliability is .90. Age-based standard scores were used for data analysis.

Teacher rating of reading—The Teacher Rating of Reading Problems* (Speece & Case, 2001; Speece et al., 2011) asked teachers to rank the overall reading ability of students using a 5-point Likert scale. Concurrent validity with standardized timed and untimed measures of reading ranged from .61 to .69. The Likert scale ratings are as follows: 5 = well above grade level, 4 = above grade level, 3 = on grade level no additional support needed, 2 = below grade level additional support needed, 1 = well below grade level intensive support needed. Raw scores were used.

Social and problem behavior rating—Teachers completed two rating scales in Fall: the Social Skills Rating System (SSRS; Gresham & Elliot, 1990) and the Strengths and

Weaknesses of ADHD Symptoms and Normal Behaviors (SWAN; Swanson et al., 2006). The subtests of the SSRS address social skills, behavior, and academic competence. Internal reliabilities were .90 for the Social Skills scale, for the Problem Behaviors scale .73 and .95 for the Academic Competence scale. Standard scores were used for data analytical purposes. The SWAN is a one-page checklist that addresses social skills, inattention and other behavior.

Data Analysis

To address the first research question we chose a strategy that would acknowledge both the individual contributions of each predictor while also recognizing that there may exist various combinations of sets of predictors equally good at predicting group membership. In order to accomplish both of these tasks, and to reduce the number of variables in the model, we performed two sets of all-subsets regressions (Miller, 2002; Speece et al., 2010); one all-subsets regression for the variables that represent initial skills and the other for the behavior rating scales. An all-subsets regression computes all possible model R^2 s for predictor set sizes from one to the total number of predictors. The model R^2 s are then rank ordered from highest to lowest within each given set size. After obtaining the highest R^2 s values for each set size, the next step is to look for a point of diminishing returns (Speece et al., 2010) where increasing the predictor set size does not substantially increase the overall amount of variance being accounted for (arbitrarily set at no more than a 2% increase in variance). At this point of diminishing returns, the best fitting models of that set size are explored for their predictive utility (that are also within 2% variance accounted for from the best model of that set). Finally, a model that includes strong initial skills predictors and strong behavioral ratings as demonstrated from the all-subsets regressions will be combined into one model that will explore these variables unique and joint predictive utility.

We also fit a Receiver Operating Characteristic (ROC) curve to estimate diagnostic decision-making using these sets of predictors to distinguish adequate and inadequate responders. A ROC curve allows for the estimation of sensitivity, specificity, and overall correct classification rate for the various cut points that could be employed from the logistic regression model. The area under the ROC curve explains the percentage of correctly identified students as adequate and inadequate responders; so that .50 is equal to chance and a 1.0 is perfect identification (generally, greater than .90 is considered excellent and .80 to .90 is considered good). Sensitivity refers to identification of true positives and is calculated by dividing the number of true positives by the sum of true positives and false negatives (i.e., students who needed intervention but were not identified as such). Further, the specificity (calculated as the true negatives divided by the true negatives and false positives) should limit the number of false positives, or students who would become good readers but are falsely identified as at-risk. Specifically, the logistic regression estimates a probability of being a responder for each person and for a pattern of scores on the four predictor variables. Jenkins, Fuchs, van den Broek, Espin, and Deno (2003) indicated that RTI procedures should have a sensitivity above 90%, thus minimizing misdirection of resources to false positives. A cut-point can be chosen for each probability value to produce classification accuracy statistics such as sensitivity, specificity, and overall classification accuracy.

To address the second research question, related to patterns of engagement, behavior and emotions observed during intervention that might explain adequate and inadequate response, we used the grounded theory approach (Glaser, 2000; Strauss & Corbin, 1998). In her role as project coordinator and participant observer, the first author led a close examination of Tier 2 and Tier 3 intervention videotapes to learn more about differences in engagement and avoidance and emotions of students found to be adequate and inadequate responders. We followed several steps related to quality indicators for qualitative research (e.g., Brantlinger, Jimenez, Klingner, Pugach, & Richardson, 2005) to strengthen confidence in the findings and to discount alternative hypotheses. Fidelity of intervention implementation was observed to be high ($M = .89$). Field notes were collected using guided theory (Denzin, 2001) to identify examples and non-examples of avoidance/engagement and achievement emotions. Behaviors were recorded in a spread sheet and three mutually exclusive categorical codes emerged: physical, verbal (and other noise making), and interaction. Next, videos were reviewed to determine which were positive or negative and to learn whether different patterns distinguished inadequate and adequate responders. Additional steps to establish trustworthiness included member checking with the interventionists, examining student assessment data, theory-checking with co-authors, and analyzing and discussing specific cases with co-authors to validate the coding process (Denzin, 2001).

Results

The first question involved the identification of variables that could predict group membership of adequate and inadequate responders to intervention. We conducted a multivariate analysis of variance and the Wilk's lambda test statistic demonstrated a statistically significant multivariate main effect of membership, $F(16, 139) = 2.191$, $p = .008$, with significant differences across most measures. Table 2 reports the means, standard deviations, and significance levels for these variables. Table 3 shows the correlations among all variables.

All-subsets regression of the initial skills variables

An all-subsets regression was conducted on the variables that represented the initial skills that the students brought to the intervention (letter word identification, word attack, comprehension, blending words, elision, fluency composite, vocabulary, teacher rating of reading performance, and verbal and non-verbal IQ). The R^2 's values for all possible subsets of all possible combinations of 10 variables were then rank ordered from highest to lowest for each given set size. Once the highest R^2 value was obtained for each given set size of 10, we looked for a point of diminishing returns, where an increase in the set-size of the predictors did not provide a sizable increase in variance accounted for (set arbitrarily at 2%). The models with variables set sizes from one to five yielded percent R^2 values of 10.2, 12.2, 15.1, 16.1, and 16.7 respectively. The remaining models of set sizes 6 through 10 yielded R^2 percent values of 17.1, 17.4, 17.7, 17.9, and 17.9 respectively. After inspecting these R^2 values, we chose the models that had 3 predictor variables. These showed an increase of 2.9 in R^2 value from the previous set-size and the next increment (a set size of 4) only yielded an increased R^2 of $16.1 - 15.1 = 1.0$. Table 4 displays the top 5 predictor models for set size of 3 predictors. What becomes clear from looking at Table 4 is that all of the top five models

contain the contain at least two of the same three variables: letter word identification, fluency composite, and blending words, and the model that contains all three explains the most variance (15.1%). The next best model only accounts for 12.9% of the variance in group membership. This as an indication that the three variables listed above are carrying the bulk of the information about the prediction of adequate and inadequate responders.

All-subsets regression of the behavioral rating variables

A second all-subsets regression was conducted on the six behavioral rating variables (SWAN hyperactivity, inattention, other and SSRS academic, problem behavior, and social). We rank ordered the R^2 's values for all possible subsets of all possible combinations of variables. The models for set sizes one through six yielded percent R^2 values of 3.2, 4.0, 4.8, 5.1, 5.2, 5.2, respectively. Because the increase in percent R^2 was only .8 between the best model of set-size one and the best model of set-size two, we chose the stay with the models that contained only one predictor; all six models appear in Table 5. This table indicates that there are two models with similar predictive utility accounting for nearly 3% of variance (SSRS Academic and SWAN Inattention).

Combining initial skill predictors and behavioral assessments

After inspecting the results of the all-subsets regression, we performed a logistic regression that includes the three initial skills predictors (Letter Word, Fluency and Blending Words) and the SSRS-Academic variable into one model. The results of this analysis appear in Table 5. The likelihood-based pseudo- R^2 for the entire model was .17, indicating that this model was able to predict 17% of the variance in group membership. An inspection of the regression weights in Table 6 indicates that three of the predictors uniquely predict group membership (Letter Word, Fluency, and Blending Words) while one predictor (SSRS-Academic) did not.

We also fit a ROC curve using data from the 156 students. The purpose of the ROC curve was to investigate if we could predict which students would be classified as responders (or nonresponders) at the end of the year using information obtained at the beginning of the year. Specifically, we took the best initial skills predictors (LetterWord, Fluency, and Blending Words) and one of the best behavioral ratings predictor (SSRS Academic) and used these four predictors to predict responders and non-responders. We chose the SRSS Academic over the SWAN Inattention because it has a slightly higher zero-order relationship with group membership, but the readers should note using either variable in the ROC curve produced a nearly identical result. A ROC curve allows for the estimation of sensitivity, specificity, and overall correct classification rate for the various cut points that could be employed from the logistic regression model. Specifically, the logistic regression estimates a probability of being a responder for each person and for their pattern of scores on the four predictor variables. A cut-point can be chosen for each probability value to produce classification accuracy statistics such as sensitivity, specificity, and overall classification accuracy. The optimal cut-point for correct classification produced a value of 87.5% correct classification, but this cut-point produced a sensitivity rate of 97.8 and a specificity of 15.0. A cut-point that optimized sensitivity and specificity produced an overall classification accuracy of 82.7%, with a sensitivity of 85.3, and a specificity of 65.0. This

could be in part because students may have been at risk, but actually responded to intervention, which is a desired outcome.

The second question addressed potential patterns of engagement, behavior, and emotions observed during intervention sessions that could distinguish adequate and inadequate responders. We used the grounded theory approach (Denzin, 2001; Glaser, 2000; Strauss & Corbin, 1998) to describe differences that were exhibited by inadequate responders in their engagement/avoidance behavior and emotions. It is important to note that not all students could be observed at all three time points depending upon their initial assignment (Dynamic or Typical) and absences during taping. By design, students in the typical condition could only receive Tier 3 after failing to respond to Tier 1 and 2; thus they would not have been observed in Fall. Of the 20 inadequate responders, 9 were observed in Fall (one was absent during taping and the remaining 10 were in Tier 1), 16 were observed in Winter (four absent), and 19 were observed in Spring (one absent).

Engagement/avoidance

The strongest and most consistent pattern we observed that contrasted adequate and inadequate responders was engagement/avoidance; three categories of engagement/avoidance emerged: physical, verbal, and interaction. Notably avoidance of all types occurred frequently during the most difficult tasks, which for inadequate responders was blending and stretching. Often times the students could not repeat their blending and stretching, even after the interventionist modeling using the instructional routine of “My turn, our turn, your turn.” Physical avoidance was defined as movement during intervention (e. g., standing, rocking, and leaving the group) and verbal avoidance (and other noise making) was defined as anything that had to do with talking or noise (e. g., yelling out answers, humming). Interaction avoidance was defined as peer-to-peer interaction or student-to-teacher interaction (e. g., arguing with teacher, talking with peer, arguing with peer, and touching a peer). Next, tapes and field notes were reexamined to discern whether these three categories included positive/engagement or negative/avoidance behaviors.

Physical

The most frequent categories of avoidance was physical, which was coded for nine, or nearly half of the inadequate responders, but was seen far less among the adequate responders. Physical movements, ordered from most to least common, included bouncing in the chair, stretching (extending arms and leaning back), sitting on knees, rubbing hands across the table or putting hand on head or face, and for some, putting their hand or shirt in their mouth. It was notable that, across observations, students demonstrated fairly consistent patterns. All of these behaviors could have been seen as negative, but on reexamination and confirmed by interventionists, at times the students still remained engaged and on task and therefore was coded as positive or engagement behavior. More frequently, however among inadequate responders, were physical behaviors that clearly demonstrated avoidance behavior and had negative outcomes. These negative physical behaviors, exhibited by the inadequate responders, included moving away from the group, going under the table, or asking frequently to go to the bathroom. Most frequently, this occurred during blending and segmenting activities, but no physical avoidance behavior occurred during dialogic reading

in fall. In winter, this avoidance behavior was exhibited during decodable reading involved refusal to pointing to words read by peers or the interventionist and similarly, if students struggled with reading independently. It is noteworthy that interventionists had been trained in positive behavior support techniques and were observed to get students back on task and engaged. It was surprising to us that there was a group of five students who were among the weakest responders (with end of year standard scores on WJ-III: Basic Reading Skills Cluster ranging from 73–79) whose behavior did not indicate they were avoiding. Masking their inadequate response, they sat quietly during intervention and participated and followed the rules, but they were passive and they copied their peers frequently.

Verbal

The second most frequent category of avoidance was verbal (or other noise producing), and as with the physical avoidance category, encompassed behaviors that could be categorized as positive/engaging or negative/avoidance. Examples of noise producing included banging fists on table, tapping pencils on table, noises produced by mouth, talking, arguing, whispering, and yelling. These negative behaviors included self-talk statements related to the difficulty of the task and not directed to the teacher such as “I can’t do this,” “I won’t do this,” “It is too hard,” “I can’t read,” “I am leaving,” “I am hungry I want a snack,” “I don’t know that word”. A small handful of inadequate responders answered when called on and appeared to be on task during the activities, but engaged in minor verbal avoidance behavior such as making deliberate noises such as exaggerated yawning. Another handful seemed to have some impulsivity issues and would blurt or shout out answers rather than raising their hand to answer questions; however, they would volunteer to read and do other activities and were engaged in the activity. This was categorized as a positive engagement behavior rather than verbal avoidance.

Interaction

The least frequent category of avoidance was interaction; this was defined as negative and positive physical and verbal interactions that included not only the student but also involvement with peers or teacher. Observed examples include talking to a peer during the intervention session, putting hands in teacher or peers’ face, putting hands on peer, volunteering to help teacher with materials, receiving high fives from teacher, and getting up to play with peer while turning in work. Statements that showed interaction with teacher involved students making eye contact and saying “I need help,” “What word is this,” “I can’t do this,” and “Help me”.

The positive behaviors that were exhibited were high fives between students and teacher, showing affirmation of students’ improvement or correct answer. The negative interaction category included teacher reprimands for inappropriate behavior such as making sounds, hitting the table, grabbing materials off table, and laughing or making statements about other students in the group (e.g. “Someone stinks in here, we need some perfume,” “Someone’s breath smells bad”). The highest frequency of negative interaction was seen for a single inadequate responder who exhibited the most avoidance behaviors during intervention. He was an outlier and seemed to need continual academic support when doing an activity and may have benefited from one to one intervention.

Emotions—The next level of coding re-examined the avoidance data through a the theoretical framework of positive and negative emotions (Pekrun et al., 2006, Pekrun et al., 2009) and patterns that emerged were largely consistent with the categories of positive emotions (hope and pride) and negative emotions (anxiety, shame, and hopelessness). *Hope* was displayed through behavior such as smiling, answering questions without being coaxed, and pointing to words (e.g., decodable reading). *Pride* was displayed by volunteering to participate in the activity by raising hand, doing independent practice on some activities, participating in activities (e.g., stretching, blending and sight word drills), and helping another student in the group with unknown words. *Hope* was also seen through phrases as “Can we do it again?” “I know” and “Can I do that?” These two emotions (hope and pride) were viewed more among the adequate responders than the inadequate responders.

The most common negative emotion that emerged was *anxiety*; this was exhibited by students by answering quietly, not answering unless someone else did first (e.g., copying other students), distancing self away from instruction or activity, not following along or pointing to words (e.g., decodable reading). There was little avoidance behavior exhibited during the sight word drills, but there was sponging (waiting for another student to respond and then follow in response) from students lacking automaticity. This was seen more in the inadequate responders than the adequate responders and appeared related to the level of difficulty of words. The students would exhibit their anxiety and frustration verbally with sighs, “Yuck,” or “I can’t do this.” Notably, across time, anxiety was exhibited more frequently, increasing from Fall to Winter from 2 to 4 inadequate responders and to 7 of the 20 inadequate responders by Spring.

Shame was the second most frequent negative emotion and was exhibited primarily by inadequate responders; often when students answering a question were corrected, refused to complete the activity, or withdrew from the group. It was also displayed verbally through phrases such as “It is too hard,” and physically when students buried their face in their arms. This emotion was not exhibited by the adequate responders. *Hopelessness* was the third most frequent negative emotion and was seen to increase across the year from one to three inadequate responders. The student who exhibited the strongest avoidance behaviors also expressed the most hopelessness through statements such as “I can’t do it,” and “I don’t get it,” “I won’t do it”, even if the interventionist was scaffolding guided practice. Again, this emotion was not exhibited by the adequate responders.

Discussion

The first research question investigated differences in initial skills that could predict adequate and inadequate response and included a relatively large set of variables related to language, phonological awareness, fluency, untimed reading, and teacher ratings of academics, behavior, and attention. Our findings suggested that letter word reading, the fluency composite and blending words explained the most variance (15%) in response among initial skills. Adding additional teacher ratings of behavior and academics, accounted for only a small amount of additional variance (3%) in group membership. Thus our findings extend and corroborate prior work on nonresponders (inadequate responders) within two-tier models (Lam & McMaster, 2013) by indicating that fluency, word identification, and

phonological awareness are important predictors of responsiveness in three-tier models. It is noteworthy that the fluency composite, was administered in less than 5 minutes and emerged as a significant predictor across all models. The salience of fluency is also underscored by study findings by Vaughn et al. (2009), who reported that second graders with oral reading fluency scores above 40 words per minute demonstrated stronger response to intervention than students with lower scores, and by O'Connor et al. (2005) who reported that all first graders who received intervention and reached a speed of 40 words per minute by the end of first grade remained on grade level for reading through third grade.

We were surprised that of the teacher ratings, only SWAN inattention and SSRS academic competence added significantly to the model. By contrast, the SSRS social skill and problem behavior subtests and the SWAN hyperactivity did not. To date, behavior scales have not been used consistently in studies to determine responsiveness and were also not significant predictors in Nelson et al.'s, (2003) literature review. Notably, however, the SSRS Academic variable was also used in Vaughn et al.'s (2009) study to examine responsiveness. In contrast, SWAN inattention, which was also statistically significant, was not seen in any prior research of RTI.

That we correctly classified 87.5% of students, yielding a sensitivity of 85.3 and a specificity of 65.0, indicates our diagnostic accuracy was better for adequate than inadequate responders. Our sensitivity value was slightly below the criteria of 90% suggested by Jenkins et al. (2003) and were between rates reported by prior researchers (95–100% range for Compton, Fuchs, Fuchs, & Bryant, 2006; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; O'Connor & Jenkins, 1999; and 65–66% reported by Speece & Case, 2001; Torgesen & Davis, 1996). Our specificity values were also lower than prior research (Compton et al: 91%, O'Connor & Jenkins: 87%, Speece & Case: 83%, Foorman et al: 63%). Differences in these rates could be due to different classification cut points used by researchers. Another explanation is that none of the prior investigations investigated responsiveness among students receiving supplemental interventions within the context of a full three-tier model, the sensitivity value could be low because students needed the intervention they received and because they responded. Relatedly, there is no prior research that fast tracked students with the weakest initial skills to Tier 3. Given that brief initial screeners and teacher rating scales predicted responsiveness, it is promising that students do not have to wait until mid-year to acquire services.

The second purpose of our study was to learn more about inadequate responders by closely observing their engagement/avoidance in contrast to adequate responders (Case et al., 2003; Morgan and Fuchs, 2007) and emotions displayed during intervention sessions (Pekrun et al., 2006, Pekrun et al., 2009). In the area of avoidance, three categories emerged: physical, verbal and interaction. The most frequent was physical avoidance, which was demonstrated across the year by nearly half of inadequate responders, but was rarely by adequate responders. However, we learned that despite moving around in their seats, some students were really engaged (seen by both adequate and inadequate responders). In contrast, there was physical avoidance that was negative, which exhibited lack of engagement in learning, and only occurred with inadequate responders. Morgan and Fuchs (2007) suggest that there is a relationship between reading skills and motivation to learn which may help to explain

some of the dynamics that were exhibited with negative avoidance behaviors. In observing the second category, verbal avoidance, we learned that behavior that might be considered negative or impulsive behavior (standing, blurting out answers, not raising hands) at times supported student engagement or revealed their excitement in knowing the answer. This finding may be consistent with teacher ratings of the SWAN Inattention; items asked teachers whether students could sustain attention on tasks or play activities, sit still, reflect on questions (control blurting out answers), and enter into conversations and games (control interrupting/intruding). It was noteworthy that avoidance occurred most often during the blending tasks, which were most difficult and potentially frustrating for students. This type of behavior was also acknowledged by Case et al. (2003), who recognized that strong teachers had the knowledge to direct that verbal impulsivity to the task at hand and not view it as negative. The third avoidance category was interaction, which was least frequent. Positive interaction was exhibited by both adequate and inadequate responders, but negative interaction was only exhibited by inadequate responders.

Next, we examined emotions displayed during intervention sessions and learned that the positive emotions of hope and pride were displayed most often by adequate responders. Interestingly, the emotion of anxiety was manifested in both adequate and inadequate responders. In contrast, among inadequate responders we saw both hopelessness and shame. Our interventionists expressed they wished they could have provided some one-to-one sessions to learn if these emotions could be changed through individualized attention and increased mastery.

Limitations and Directions for Future Research

As with any school based research the study had several limitations, but also some directions for future research. First, the sample size of 20 inadequate responders was small, thus we were unable to explore whether the intervention condition differentially influenced inadequate response. Nor did any subgroup profiles emerge among the groups. Second, the larger study used two RTI models, which led to different amounts of intervention (dosage) being confounded with condition. Third, intervention was given by trained researchers; additional research is needed with other school based personnel to improve generalizability of findings. Additional research is also needed with students with different backgrounds. Fourth, in the larger study a standard protocol was followed so it is unclear if different findings would occur using a more individualized problem solving approach. Fifth, although teacher checklists of behavior slightly improved prediction of inadequate response, further observational research is needed to confirm the importance of task avoidance behaviors and emotions. Sixth, the observers of the intervention was not blind to condition. Finally, given that not responding within the context of intervention was defined as being at or below the 40th percentile, the cut point for the outcome analysis was the 25th percentile. Thus, a student could have appeared unresponsive (between the 25th and 40th percentile) during intervention, but might have achieved adequate response (above the 25th percentile).

Implications for Practice

The limitations notwithstanding, findings from the present study have several important implications for improving multi-tier RTI implementation. First, it is possible to predict

responsiveness fairly accurately from beginning of first grade fluency screeners (letter sound, sight word, and phonemic decoding fluency) combined with letter word identification and blending tasks. In the present study, we were less concerned about false positives, because we wanted to identify students at risk, and to mitigate that risk through early intervention. Thus, using constructs that have been supported in research for over a decade, students can receive the most intensive interventions without having to wait-to-fail in Tier 1 or 2, as failure and frustration could contribute to negative emotions and avoidance behaviors. Furthermore, study findings lead credence to teacher judgment about students' academic performance and inattention, although they did not add a substantial amount of prediction beyond student scores. A second implication is that our findings confirm and extend the literature indicating that some students will require additional intervention due to inadequate response, even to three-tier RTI models, and even if fast-tracked to Tier 3. A third and related implication is that, as a field, we need more research about how to tailor interventions to ensure they are robust enough to help students who are persistently inadequate responders. Our exploration of engagement/avoidance and emotions displayed by students suggests that it may be particularly important for interventionists to tune in to emotions like hopelessness and shame and to carefully consider why students may avoid tasks in order to increase motivation and engagement. It also may be possible to use data based regrouping to create smaller groups, or some one-to-one intervention sessions, to ensure tasks are not too difficult and that pacing is slow enough to ensure mastery, while not initiating boredom.

Acknowledgments

This work was supported by (a) a Multidisciplinary Learning Disabilities Center Grant P50HD052120 from the National Institute of Child Health and Human Development

References

- Al Otaiba S, Fuchs D. Characteristics of children who are unresponsive to early literacy intervention. A review of the literature. *Remedial and Special Education*. 2002; 23(5):300–316.
- Al Otaiba S, Connor CM, Folsom JS, Greulich L, Wanzek J, Schatschneider C, Wagner RK. To wait or to intervene immediately: A randomized experiment examining first grade response to intervention in reading. Manuscript in press in *Exceptional Children*. (in press).
- Bereiter, C.; Brown, A.; Campione, J.; Carruthers, I.; Case, R.; Hirshberg, J.; Treadway, GH. *Open court reading*. Columbus, OH: SRA McGraw-Hill; 2002.
- Berkeley S, Bender WN, Peaster L, Saunders L. Implementation of Response to Intervention: A snapshot of progress. *Journal Of Learning Disabilities*. 2009; 42(1):85–95. [PubMed: 19103800]
- Brantlinger E, Jimenez R, Klingner J, Pugach M, Richardson V. Qualitative studies in special education. *Exceptional Children*. 2005; 71:195–207.
- Burns MK, Appleton JJ, Stehouwer JD. Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment*. 2005; 23(4):381–394. Retrieved from <http://search.proquest.com/docview/61913319?accountid=4840>.
- Case LP, Speece DL, Molloy DE. The validity of a response-to- instruction paradigm to identify reading disabilities: A longitudinal analysis of individual differences and contextual factors. *School Psychology Review*. 2003; 32(4):557–582.
- Compton DL, Fuchs D, Fuchs LS, Bryant JD. Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*. 2006; 98(2):394–409.

- Compton DL, Gilbert JK, Jenkins JR, Fuchs D, Fuchs LS, Cho E, Barquero LA, Bouton B. Accelerating chronically unresponsive children to tier3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities*. 2012; 45:204–216. doi: <http://dx.doi.org/10.1177/0022219412442151>. [PubMed: 22491810]
- Cone, JD. "Idiographic, nomothetic, and related perspectives in behavioral assessment". In: Nelson, RO.; Hayes, SC., editors. *Conceptual foundations of behavioral assessment*. New York: Guilford; 1986. p. 111-128.
- 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]
- Denzin, N. Strategies of multiple triangulation. In: Conrad, CF.; Haworth, JG.; Lattuca, LR., editors. *Qualitative research in higher education: Expanding perspectives*. 2nd ed. Boston, MA: Pearson; 2001. p. 317-327.
- 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.
- Fuchs D, Fuchs LS, Compton DL. Smart RTI: A Next-Generation Approach to Multilevel Prevention. *Exceptional Children*. 2012; 78(3):263–279. [PubMed: 22736805]
- Fuchs D, Fuchs LS, Stecker PM. The "Blurring" of Special Education in a New Continuum of General Education Placements and Services. *Exceptional Children*. 2010; 76(3):301–323.
- Fuchs D, Fuchs LS, Thompson A, Al Otaiba S, Yen L, Yang NJ, Braun M, O'Connor RE. Exploring the Importance of Reading Programs for Kindergartners with Disabilities in Mainstream Classrooms. *Exceptional Children*. 2002; 68(3):295–311.
- Fuchs D, Mock D, Morgan PL, Young CL. Responsiveness-to-Intervention: Definitions, Evidence, and Implications for the Learning Disabilities Construct. *Learning Disabilities Research & Practice* (Blackwell Publishing Limited). 2003; 18(3):157–171.
- Fuchs LS, Vaughn S. Responsiveness-to-Intervention: A Decade Later. *Journal of Learning Disabilities*. 2012; 45(3):195–203. [PubMed: 22539056]
- Gambrell LB, Palmer B, Codling R, Mazzone S. Assessing motivation to read. *Reading Teacher*. 1996; 49(7):518. doi: <http://dx.doi.org/10.1598/RT.49.7.2>.
- 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/rrq45.24>.
- Glaser BG. The future grounded theory. *Grounded Theory Review*. 2000; 2:1–18. doi: <http://dx.doi.org/10.1177/104973299129122199>.
- Good, RH.; Kaminski, RA. *Dynamic indicators of basic literacy skills*. 6th ed.. Eugene, OR: University of Oregon, Institute for the Development of Educational Achievement; 2001.
- Gresham, FM.; Elliot, SN. *Social skills rating system*. Circle Pines, MN: American Guidance Service; 1990.
- 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.
- Jenkins JR, Fuchs LS, van d, Broek, Espin C, Deno SL. Accuracy and fluency in list and context reading of skilled and RD groups: Absolute and relative performance levels. *Learning Disabilities Research & Practice*. 2003; 18(4):237–245.

- Kaufman, A.; Kaufman, N. Kaufman brief intelligence test (2nd ed.). Circle Pines, MN: American guidance Service Publishing; 2004.
- 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](http://dx.doi.org/10.1016/S0885-2006(99)80038-6).
- 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.
- 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
- McAlenney AL, Coyne MD. Identifying at-risk students for early reading intervention: Challenges and possible solutions. *Reading and Writing Quarterly*. 2011; 27:306–323. doi: <http://dx.doi.org/10.1080/10573569.2011.596100>.
- Mellard D, McKnight M, Jordan J. RtI tier structures and instructional intensity. *Learning Disabilities Research & Practice*. 2010; 25:217–225.
- Mellard DF, McKnight M, Woods K. Response to Intervention Screening and Progress-Monitoring Practices in 41 Local Schools. *Learning Disabilities Research & Practice* (Blackwell Publishing Limited). 2009; 24(4):186–195.
- Miller, AJ. Subset Selection In Regression. 2nd ed.. New York: Chapman & Hall; 2002.
- Morgan P, Fuchs D. Is there a bidirectional relationship between children's reading skills and children's motivation? *Exceptional Children*. 2007; 73(2):165–183.
- Nelson J, Benner GJ, Gonzalez J. Learner Characteristics that Influence the Treatment Effectiveness of Early Literacy Interventions: A Meta-Analytic Review. *Learning Disabilities Research & Practice* (Blackwell Publishing Limited). 2003; 18(4):255.
- 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. [PubMed: 16392695]
- O'Connor RE, Jenkins JR. Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*. 1999; 3(2):159–197. doi: http://dx.doi.org/10.1207/s1532799xssr0302_4.
- Onatsu-Arivilommi T, Nurmi J-E. The role of task-avoidant and task-focussed behaviors in the development of reading and mathematical skills during the first schoolyear: A cross-lagged longitudinal study. *Journal of Educational Psychology*. 2000; 92:478–491. doi: <http://dx.doi.org/10.1037//0022-0663.92.3.478>.
- Pekrun R, Elliot AJ, Maier MA. Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*. 2006; 98(3):583–597.
- Pekrun R, Elliot AJ, Maier MA. Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology*. 2009; 101(1): 115–135.
- Quirk M, Schwanenflugel P. Do supplemental remedial reading programs address the motivational issues of struggling readers? An analysis of five popular programs. *Reading Research and Instruction*. 2004; 43(3):1–19. doi: <http://dx.doi.org/10.1080/19388070509558408>. [PubMed: 20076771]
- Shinn, MM.; Shinn, MR. AIMSweb. Eden Prairie, MN: Edformation; 2004.
- Speece DL, Case LP. Classification in the context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology*. 2001; 93(4):735–749.
- Speece DL, Ritchey KD, Silverman R, Schatschneider C, Walker CY, Andrusik KN. Identifying children in middle childhood who are at risk for reading problems. *School Psychology Review*. 2010; 39:258–276. [PubMed: 21472039]
- 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]
- Strauss, A.; Corbin, J. Basics of qualitative research: Techniques and procedures for developing grounded theory. 2d ed.. Thousand Oaks, CA: Sage; 1998.

- Swanson, J.; Schuck, S.; Mann, M.; Carlson, C.; Hartman, K.; Sergeant, J.; Clevenger, W.; Wasdell, M. Categorical and dimensional definitions and evaluations of symptoms of ADHD; the SNAP and SWAN rating scales. Irvine: University of California; 2006.
- 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]
- Torgesen JK, Davis C. Individual differences variables that predict response to training in phonological awareness. *Journal Of Experimental Child Psychology*. 1996; 63(1):1. doi: <http://dx.doi.org/10.1006/jecp.1996.0040>. [PubMed: 8812025]
- Torgesen, JK.; Wagner, RK.; Rashotte, CA. Test of word reading efficiency. Austin, TX: PRO-ED; 1997.
- Vaughn S, Linan-Thompson S, Hickman P. Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children*. 2003; 69(4):391–409.
- 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*. New York, NY: US: Springer Science + Business Media; 2007. p. 234–243.
- 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 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: 16583795]
- 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.
- Wagner, RK.; Torgesen, JK.; Rashotte, CA. Comprehensive test of phonological processing. Austin, TX: PRO-ED; 1999.
- Wanzek J, Vaughn S. Research-Based Implications From Extensive Early Reading Interventions. *School Psychology Review*. 2007; 36(4):541–561.
- Wigfield A, Guthrie J. Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology*. 1997; 89(3):420–432. doi: <http://dx.doi.org/10.1037//0022-0663.89.3.420>.
- Woodcock, R.; McGrew, K.; Mather, N. Woodcock Johnson-III tests of achievement. Itasca, IL: Riverside; 2001.
- Zirkel PA, Thomas LB. State Laws and Guidelines for Implementing RTI. *Teaching Exceptional Children*. 2010; 43(1):60–73.

Table 1

Tiers Received by Response Status

	Response Status	
	Adequate	Inadequate
112	1	1
121	1	0
122	56	3
123	10	6
133	3	0
221	24	0
222	12	3
223	7	2
233	4	1
332	4	0
333	14	4
Total	136	20

Note: Tier 112 indicates students in Tier 1, Tier 1, and Tier 2.

Table 2

Means and Standard Deviations for Initial Skills and Child Characteristics

Construct	Adequate	Inadequate	Sig.
	M (SD)	M (SD)	
Initial Skills			
Blending ^S	11.70 (2.3)	9.35 (2.7)	.000
Elision ^S	8.21 (2.6)	6.05 (2.4)	.001
Picture Vocabulary [*]	478.00 (12)	475.00 (14)	.304
Letter Word [*]	392.00 (19)	383.00 (15)	.047
Word Attack [*]	450.00 (20)	439.00 (24)	.027
Passage Comp. [*]	421.00 (18)	411.00 (17)	.027
Fluency Composite	.301 (3.8)	−3.00 (2.7)	.000
Speece Rating ^R	2.04(.85)	1.60 (.60)	.026
Verbal IQ	93.20 (14)	85.40 (9.2)	.018
Non-verbal IQ	90.60 (12.8)	82.90 (7.8)	.010
SSRS Social Skills ^S	99.70 (16)	98.70 (16)	.794
SSRS Behavior Problems ^S	96.20 (12)	97.30 (14)	.728
SSRS Academic ^S	88.60 (11)	82.70 (12)	.026
SWAN Inattention ^S	31.20 (9.8)	26.10 (9.8)	.033
SWAN Hyperactivity ^S	35.30 (9.5)	32.80 (7.9)	.283
SWAN Other ^S	50.20 (14)	49.00 (14)	.697

Note.

* = W-score,

S = Standard Score,

R = Raw Score.

Initial Skills = Blending (CTOPP; Wagner et al., 1999); Elision (CTOPP; Wagner et al., 1999); Picture Vocabulary (Woodcock et al., 2001); Letter Word Identification (Woodcock et al., 2001); Word Attack (Woodcock et al., 2001); Passage Comprehension (Woodcock et al.); Fluency Composite: Word Identification Fluency (WIF; Fuchs et al., 2004b), Letter Sound Fluency (AIMSweb; Shinn & Shinn, 2004), Oral Reading Fluency (DIBELS; Good & Kaminski, 2001), Sight Word Efficiency (TOWRE; Torgesen et al., 1997b), Phonemic Decoding Efficiency (TOWRE; Torgesen et al., 1997b); Speece Rating Scale (Speece & Case, 2001); Verbal and Non-verbal IQ (Kaufman Brief IQ; Kaufmann & Kaufmann, 2004); Social Skills, Behavior Problems, Academic (SSRS; Gresham & Elliot, 1990); Inattention, Hyperactivity, Other (SWAN; Swanson et al., 2006).

Table 3

Correlations among Constructs

Constructs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<i>Blending</i> ^S	1.00															
<i>Elision</i> ^S	0.57	1.00														
<i>Picture Vocab</i> [*]	0.36	0.45	1.00													
<i>Letter Word</i> [*]	0.55	0.45	0.36	1.00												
<i>Word Attack</i> [*]	0.64	0.46	0.43	0.55	1.00											
<i>Passage Comp</i> [*]	0.32	0.32	0.3	0.56	0.35	1.00										
<i>Fluency Composite</i>	0.54	0.49	0.33	0.82	0.56	0.57	1.00									
<i>Speech Rating</i> ^R	0.39	0.43	0.34	0.56	0.44	0.39	0.50	1.00								
<i>Verbal IQ</i> ^S	0.42	0.45	0.61	0.37	0.33	0.37	0.39	0.33	1.00							
<i>Non-verbal IQ</i> ^S	0.30	0.40	0.24	0.27	0.31	0.23	0.30	0.25	0.45	1.00						
<i>SSRS Social Skills</i> ^S	0.16	0.14	0.29	0.16	0.19	0.13	0.16	0.24	0.28	0.13	1.00					
<i>SSRS Problem Behavior</i> ^S	-0.01	0.01	-0.03	0.07	0.00	0.03	0.07	-0.04	-0.17	-0.09	-0.57	1.00				
<i>SSRS Academic</i> ^S	0.33	0.31	0.48	0.43	0.38	0.34	0.43	0.44	0.48	0.31	0.50	-0.29	1.00			
<i>SWAN Inattention</i> ^S	0.34	0.37	0.42	0.39	0.37	0.34	0.47	0.55	0.46	0.30	0.56	-0.35	0.56	1.00		
<i>SWAN Hyperactivity</i> ^S	0.28	0.25	0.30	0.2	0.24	0.23	0.24	0.22	0.38	0.21	0.48	-0.49	0.48	0.67	1.00	
<i>SWAN Other</i> ^S	0.25	0.19	0.26	0.15	0.18	0.17	0.20	0.22	0.3	0.12	0.66	-0.59	0.66	0.65	0.74	1.00

Note.

* = W-score.

S = Standard Score.

R = Raw Score.

Initial Skills = Blending (CTOPP; Wagner et al., 1999); Elision (CTOPP; Wagner et al., 1999); Picture Vocabulary (Woodcock et al., 2001); Letter Word Identification (Woodcock et al., 2001); Word Attack (Woodcock et al., 2001); Passage Comprehension (Woodcock et al., 1999); Fluency Composite: Word Identification Fluency (WIF; Fuchs et al., 2004b), Letter Sound Fluency (AIMSweb; Shinn & Shinn, 2004), Oral Reading Fluency (DIBELS; Good & Kaminski, 2001), Sight Word Efficiency (TOWRE; Torgesen et al., 1997b), Phonemic Decoding Efficiency (TOWRE; Torgesen et al., 1997b); Speech Rating Scale (Speech & Case, 2001); Verbal and Non-verbal IQ (Kaufman Brief IQ; Kaufmann & Kaufmann, 2004); Social Skills, Behavior Problems, Academic (SSRS; Gresham & Elliot, 1990); Inattention, Hyperactivity, Other (SWAN; Swanson et al., 2006).

Table 4

R^2 values for the top five initial skill models with a predictor set size = 3

Predictors	R^2
Letter Word [*] , Fluency, Blending ^S	.151
IQ-non-verbal ^S , Fluency, Blending ^S	.129
Word Attack [*] , Fluency, Blending ^S	.128
Fluency, Elision ^S , Blending ^S	.126
Letter Word [*] , Fluency, Elision ^S	.125

Note.

* W-score,

S = Standard Score,

R = Raw Score

Initial Skills = Letter Word Identification (WCJ-III; Woodcock et al., 2001); Fluency Composite: Word Identification Fluency (WIF; Fuchs et al., 2004b), Letter Sound Fluency (AIMSweb; Shinn & Shinn, 2004), Oral Reading Fluency (DIBELS; Good & Kaminski, 2001), Sight Word Efficiency (TOWRE; Torgesen et al., 1997b), Phonemic Decoding Efficiency (TOWRE; Torgesen et al., 1997b); Blending (CTOPP; Wagner et al., 1999); IQ non-verbal (KBIT; Kaufman & Kaufman, 2004); Word Attack (WCJ-III; Woodcock et al., 2001); Elision (CTOPP; Wagner et al., 1999).

Table 5

R^2 values for the six behavior rating models with a predictor set size = 1

Predictors	R^2
SSRS-Academic ^S	.032
SWAN-Inattention ^S	.029
SWAN-Hyperactivity ^S	.008
SWAN-Other ^S	.001
SSRS-Problem Behavior ^S	.001
SSRS-Social ^S	.0004

Note.

S = Standard Scores

Social Skills Rating Scale (SSRS; Gresham & Elliot, 2004), Strengths and Weaknesses of ADHD Symptoms and Normal Behaviors (SWAN; Swanson et al., 2006)

Table 6

Results from a Logistic Regression Predicting Response to Intervention

Parameters	Estimate	SE	p value	Odds Ratio	95% CI - Lower Bound	95% CI - Upper Bound
Intercept	24.19	11.15				
Letter Word*	-0.07	0.029	0.0172	0.93	0.881	0.988
Fluency	0.5	0.169	0.003	1.65	1.187	2.302
Blending ^S	0.34	0.1367	0.0124	1.41	1.077	1.84
SSRS Academic ^S	0.02	0.028	0.4764	1.02	0.966	1.078

Note.

* W-score,

S = Standard Score,

R = Raw Score.

Initial Skills = Letter Word Identification (WJC-III; Woodcock et al., 2001); Fluency Composite: Word Identification Fluency (WIF; Fuchs et al., 2004b), Letter Sound Fluency (AIMSWeb; Shinn & Shinn, 2004), Oral Reading Fluency (Good & Kaminski, 2001), Sight Word Efficiency (TOWRE; Torgesen et al., 1997b), Phonemic Decoding Efficiency (TOWRE; Torgesen et al., 1997b); Blending (CTOPP; Wagner et al., 1999); SSRS Academic (Social Skills Rating Scale (Gresham & Elliot, 1990)).