Special Series Article

Hammill Institute on Disabilities

Learning Disability Quarterly
2014, Vol. 37(4) 192–203

DOI: 10.1177/0731948713518336
ldq.sagepub.com
SAGE

Understanding Unresponsiveness to Tier 2 Reading Intervention: Exploring the Classification and Profiles of Adequate and Inadequate Responders in First Grade

Jessica R. Toste, PhD[1], Donald L. Compton, PhD[2], Douglas Fuchs, PhD[2], Lynn S. Fuchs, PhD[2], Jennifer K. Gilbert, PhD[2], Eunsoo Cho, MEd[2], Laura A. Barquero, MS[2], and Bobette D. Bouton, MEd[2]

Abstract

The purpose of the current study was to examine academic and cognitive profiles of first graders who responded adequately and inadequately to intensive small-group reading intervention (Tier 2), as well as assess how these profiles differ based on the criteria used for classification of unresponsiveness. Nonresponders were identified using two different methods: (a) reading composite with weighted standardized scores for untimed word identification and word attack, timed sight word reading and decoding, and reading comprehension at the end of first grade (n = 23; 18.4%), and (b) local norms on first grade word identification fluency (WIF; n = 31; 24.8%). Repeated measures ANOVAs were used to assess the difference between responders and nonresponders on four separate profiles (i.e., academic and cognitive profiles, with groups identified using reading composite and WIF criteria for unresponsiveness). Significant level effects were found using the first-grade reading composite and the WIF criteria, indicating that the groups differ from one another across domains. Interestingly, there were only significant shape effects found when using the WIF criteria, suggesting relative strengths and weaknesses distinguish the groups. These findings suggest potentially important considerations related to identification and placement of students in appropriately intensive and targeted interventions.

Keywords

nonresponders, early reading, intensive intervention

The ability to read is an important skill for children's academic success and overall well-being (National Reading Panel, 2000; Snow, Burns, & Griffin, 1998); however, this is not an easy task for all students. Researchers suggest that at least 20% of children have some difficulty in mastering the skills necessary to become proficient readers (Fletcher, Lyon, Fuchs, & Barnes, 2007; Lyon, 1995). Longitudinal research confirms that many of these early reading problems persist, and that students who struggle with learning to read in the primary grades are likely to struggle with reading throughout their school years (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Hall & Moats, 1999; Vaughn, Linan-Thompson, & Hickman, 2003). There is substantial evidence to suggest that early identification of students who are at-risk of reading difficulties and subsequent intervention can enhance the likelihood of positive learning outcomes. Response to intervention (RTI) builds on this logic by shifting the focus in delivery of school services toward assessing risk and providing targeted instruction for students at high risk for poor reading outcomes (D. Fuchs, Fuchs, & Compton, 2012).

The RTI framework is a multitiered system that is typically organized within three tiers of service delivery. Each tier represents a continuum of prevention and intervention, with the nature of the intervention becoming more intensive at each tier (Vaughn & Fuchs, 2003). We recognize that there are varying conceptualizations of how multitiered models are applied in schools. For the purposes of this study, we define a multitiered model as beginning with high-quality classroom instruction and universal screening (Tier 1), leading to more targeted intervention and assessment for struggling students (Tier 2), and moving to intensive support services and assessment for students who have not responded adequately to previous intervention (Tier 3).

[1]The University of Texas at Austin, USA
[2]Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Jessica R. Toste, Department of Special Education, The University of Texas at Austin, 1 University Station D5300, Austin, TX 78712, USA.
Email: jessicarosetoste@gmail.com

Over the past decade, RTI has been adopted as a school service delivery model in many school districts across the country. Recent research has shown that a considerable minority of students show inadequate improvement in reading performance even with intensive, evidence-based instructional approaches (Al Otaiba & Fuchs, 2002; D. Fuchs et al., 2012; Mathes, Howard, Allen, & Fuchs, 1998; Torgesen, 2000). Although they share the goal of providing more efficient and effective support for struggling readers, RTI systems may differ in a variety of ways—this includes the number of tiers, the types of interventions being used, and the personnel who deliver the interventions (Fuchs & Deshler, 2004). As such, there are a number of questions that remain unanswered yet are of utmost concern to researchers and practitioners alike. For instance, are there specific skills that are consistently reported as difficulties for nonresponders? How do responders and nonresponders differ? What criteria should be used to determine whether a student is an adequate responder? We look to the literature to answer the first question. Then, using data from the study, we attempt to answer the latter two.

Characteristics of Inadequate Responders

Within a RTI framework, students who experience difficulty or do not demonstrate the expected growth in their reading skills when receiving high-quality classroom instruction (Tier 1) are provided with more intensive support. Once Tier 2 instruction is delivered, practitioners evaluate student responsiveness and, more specifically, students must be categorized as adequate or inadequate responders. When this Tier 2 instruction is provided by researchers, estimates of nonresponders range from 2% to 6% of the general school-age population (Torgesen, 2000). An understanding of the characteristics and profiles of students who are unresponsive to interventions can help improve instruction as well as choose screening measures to effectively select students to receive Tier 2 and 3 interventions.

In the past, researchers had posited that students who are unresponsive to intervention have difficulties with phonological discrimination (Tallal, 1990), phonological awareness skills (Torgesen, Wagner, & Rashotte, 1994), naming speed (Wolf, 1991; Wolf & Bowers, 1999), cognitive or language skills (Bishop & Adams, 1990; Menyuk & Chesnick, 1997), and attention or behavior problems (Ackerman, Dykman, & Gardner, 1990). Al Otaiba and Fuchs (2002) conducted a review of the literature, summarizing 23 studies of early literacy interventions. They found that poor phonological awareness characterized the majority of unresponsive students, in addition to several other characteristics that were associated with unresponsiveness: phonological memory, orthographic processing, rapid naming, intelligence, and attention or behavior. In their meta-analysis, Nelson, Benner, and Gonzalez (2003) reviewed 30 studies of early literacy interventions. In examining differences between students who were adequate and inadequate responders to intervention, they reported moderate effect sizes for rapid naming (0.51), problem behavior (0.46), phonological awareness (0.42), and smaller effect sizes for letter knowledge (0.35), memory (0.31), and IQ (0.26).

In a longitudinal investigation of responsiveness to intervention, Al Otaiba and Fuchs (2006) examined student characteristics that reliably predicted adequate and inadequate response to generally effective early literacy interventions. Consistent with previous research, unresponsive students scored significantly lower than students who were responsive in assessments of vocabulary, rapid naming, problem behavior, and verbal memory. Fletcher and colleagues (2011) examined the cognitive attributes of first-grade students grouped as adequate responders, inadequate responders, and typically developing based on decoding and fluency criteria. Phonological awareness was found to be the most significant contributor to group differentiation; with measures of rapid naming, syntactic comprehension/working memory, and vocabulary also found to distinguish adequate and inadequate responders.

In sum, it is clear that there are characteristics that distinguish students who are adequate responders from those who are inadequate responders. These characteristics represent a range of academic and cognitive skills. It would be of interest to separately examine academic and cognitive profiles to develop a more complete understanding of the differences between responders and nonresponders.

Criteria for Determining Inadequate Response

One challenge that implementers of RTI must face is defining inadequate response or *unresponsiveness*. Practitioners' decisions about who requires additional support should be guided by empirical evidence (D. Fuchs, Fuchs, & Vaughn, 2008). However, there is variability in what is considered unresponsive and how this is defined, with little agreement on what constitutes adequate response to intervention (McMaster, Fuchs, Fuchs, & Compton, 2005; O'Connor & Klingner, 2010). Different methods for defining unresponsiveness often result in having different groups of students moving to more intensive (Tier 3) intervention and/or being identified as having reading disabilities.

Researchers have used multiple methods for determining unresponsiveness. For example, Torgesen and his colleagues (2001) assessed students at the end of their tutoring intervention and computed standard scores on the Woodcock Reading Mastery Tests. Those scoring below the 25th percentile were determined to be unresponsive, whereas those at or above the 25th percentile were considered responsive.

A similar approach was taken by Good, Simmons, and Kame´enui (2001) by measuring performance at the end of a third grade tutoring program. A raw score on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) oral reading fluency measure was compared against a criterion-referenced benchmark associated with performance on state testing. Students with final DIBELS scores below 40 were considered unresponsive, and those with scores at or above 40 were considered responsive.

Speece and Case (2001) utilized a dual discrepancy method in which they considered students' level and rate (slope) of performance in examining the difficulties in reading, reading-related, and behavioral measures. Similarly, L. S. Fuchs, Fuchs, and Compton (2004) utilized a method in which they identified responders and nonresponders based on a slope criterion and a posttreatment level of performance. Progress monitoring data were collected for each student during tutoring and a slope of academic performance was calculated. For both of these measures, a normative cut-point was set and students who performed below this cut-point were classified as unresponsive, whereas the others were classified as responsive. Although these studies suggest various methods that can be used to classify adequate response to intervention, it is less clear whether different RTI identification methods result in the same groups of responsive and nonresponsive students and if these groups have differing profiles.

Research Objectives

The purpose of the current study was to further our understanding of unresponsiveness to Tier 2 intervention, with the ultimate goal of supporting researchers and practitioners in making informed decisions related to the identification of students whose needs would not be met using current best practices in Tiers 1 and 2. The objectives of the study were twofold in that we sought to (1) examine academic and cognitive profiles of first graders who responded adequately and inadequately to intensive small-group reading intervention (Tier 2), and (2) assess how these profiles differ based on the criteria used for classification of unresponsiveness.

Method

Participants and Procedures

Participants were drawn from a federally funded study examining the efficacy of RTI as a means to prevent and identify students with RD (Compton, Fuchs, Fuchs, & Bryant, 2006). Participants were selected from 1,336 students in three cohorts (during consecutive years) in 73 first-grade classrooms in 15 schools in urban and suburban districts located in Nashville. The larger study, as described by Gilbert and colleagues (2013), included three cohorts of students. The Year 1 cohort (n = 712) represented a developmental sample of students followed from the fall of first grade through the end of fourth grade. We used first-grade word identification fluency (WIF) progress monitoring data and end-of-second-grade RD status to estimate WIF performance cut-points (i.e., level and slope) associated with end-of-second-grade RD. These WIF cut-points were used to categorize students as unresponsive to Tier 2 intervention in the Years 2 and 3 cohorts. The Years 2 and 3 cohorts (n = 624) were first-grade students involved in a randomized control trial (RCT) examining the efficacy of 14 weeks of supplementary intervention for students identified as unresponsive to Tier 1 instruction (Gilbert et al., 2013). The procedures for each cohort are described below, followed by a description of the first graders from the Year 2 and 3 cohorts who represent the sample of the current study.

Year 1 cohort. In mid-September, we assessed every consented child with three 1-minute study identification measures: WIF-screen, rapid letter naming (RLN), and rapid sound naming (RSN). We used a factor score comprising three measures to divide the 712 students into high-, average-, and low-performing groups and then randomly selected study students from each group. We oversampled low-performing students to increase the number of struggling readers in the developmental sample. A total of 485 students were included: 310 low-study-entry, 83 average-study-entry, and 92 high-study-entry. Participant selection occurred in late September and early October of first grade. Students were then administered a pretest battery, progress monitored using WIF for 17 weeks, posttested at the end of first grade, and followed-up at the end of second grade on a standardized battery of reading measures. At follow-up in spring of second grade, 130 of the original 485 students (27% of the sample) had moved from the district and were unavailable for assessment, leaving 355 students with complete data across first and second grade (for details see Compton et al., 2010). The Little's *Missing Completely at Random* (MCAR) test indicates that the data were missing at random (i.e., no identifiable pattern exists to the missing data), $\chi^2(71) = 70.737$, $p = .486$, and therefore, cases with incomplete data can be dropped from the analyses without creating bias. The 355 students with complete data represent the independent sample on which first-grade WIF growth parameters with an associated distal (end-of-second grade) outcome of RD were determined.

As in Compton et al. (2006; Compton et al., 2010), we classified students as RD at the end of second grade using a composite that summed weighted standardized scores for untimed word identification and word attack, timed sight word reading and decoding, and reading comprehension. The distinction between accuracy and rate of word identification was recognized through inclusion of untimed and timed measures (see Biemiller, 1977; Lovett, 1984). The weighting factor for each word identification and decoding measure was .167; for comprehension, the weighting factor was .333. This allowed the composite measure of reading to be equally weighted across untimed word-level reading and decoding, timed word-level reading and decoding, and reading comprehension. The composite was used to provide a balanced representation of reading ability by limiting the effects of a single reading skill on the classification. Students with scores below the 16th percentile on the composite were classified RD. In this way, 54 students (15% of the sample) were identified as RD at the end of second grade. To identify WIF cut-points associated with later RD, we placed a 95% confidence interval around WIF slope and final intercept for the 54 RD students and selected the upper limit of each for a locally normed cut-point (intercept = 21.41 words; slope = 0.67 words/week). Both WIF growth parameter cut-points were then used to identify students in Year 2 and 3 cohorts who were unresponsive to Tier 2 intervention.

Year 2 and 3 cohorts. In mid-September, we asked first-grade teachers to nominate students in the lower half of their class in reading. We assessed every nominated child who was consented (n = 624), with the same three 1-minute study identification measures. We then used a factor score to divide the 624 students into high-, average-, and low-performing groups and retained students from the low group for study inclusion. A total of 438 students were identified as initially low performing. Then, beginning the first week of October, we administered WIF short-term progress monitoring (PM) for 6 weeks, each time with an alternate form; during this 6-week period, students received regular reading instruction in their classrooms (Tier 1). At the conclusion of short-term PM, 10 (2.28%) of the students had moved from the district and were unavailable for assessment. One additional student was removed from the study due to scheduling difficulties.

Following short-term PM, growth modeling was used to select the students who were unresponsive to Tier 1. Because no agreed on definition of response is available in the literature, we designated unresponsiveness to Tier 1 by rank-ordering students on 6-week WIF intercept and slope, and then selecting roughly the bottom half of the 427 students (232 students or 54.33% of the at risk sample). Responders performed better on 6-week intercept, $F(1,426) = 214.96$, $p < .0001$, $d = 1.99$ (responder M = 27.83, SD = 8.86; nonresponders M = 13.65, SD = 5.22), and slope, $F(1,426) = 116.77$, $p < .0001$, $d = 1.52$, per week (responder M = 1.99, SD = 0.76; nonresponders M = 1.09, SD = 0.40). In mid-November, the 232 students deemed unresponsive to Tier 1 instruction received a battery of standardized tests, administered individually by trained examiners (each of whom had demonstrated at least 90% accuracy during practice assessments). The battery comprised measures of phonemic awareness, rapid naming, oral vocabulary, listening comprehension, untimed word identification skill, untimed decoding skill, sight word efficiency, and phonemic decoding efficiency. During testing, three students moved, and one was removed due to refusal to participate.

Students who completed the pretest battery (n = 228) were then randomly assigned (in a 2:1 ratio) to treatment (n = 149) or control (n = 79). Teachers were asked to complete an attention rating scale (SWAN; Swanson et al., 2004) for each student regardless of treatment/control status. Students received tutoring from trained research assistants in small groups (two to three students per group) for 45 minutes three times per week in addition to their classroom reading instruction. Treatment was considered Tier 2 because it comprised scripted, supplemental tutoring program that focused on phonological awareness, sight words, letter sounds, decoding, and reading fluency (see Gilbert et al., 2013, for a full description of tutoring). Students assigned to control continued to receive Tier 1 reading instruction in their classrooms. Because the purpose of this article was to examine response to Tier 2 intervention, control students were eliminated from further analyses. Tutoring began the second week of December. Throughout the 14 weeks of tutoring, students completed weekly WIF assessments. Tutors also completed the SWAN rating scale when tutoring ceased.

Current study sample. A total of 125 students (48% female) who were enrolled in Tier 2 had complete data and are included in the current sample. Forty-five percent of students were White, 41% Black, 8% Hispanic, and 6% of students belonged to other ethnic groups. Sixty-six percent (n = 82) of students qualified for free and reduced lunch, and 14% (n = 17) had an individual education plan.

Classification of Responders and Nonresponders

Unresponsiveness to Tier 2 intervention was classified in two ways for the purposes of examining whether profiles of responders and nonresponders would differ depending on how performance was operationalized. The sample of 125 students was first classified as responders or nonresponders based on a composite score of reading performance at the end of first grade. Summed weighted standardized scores for untimed word identification and word attack, timed sight word reading and decoding, and reading comprehension at the end of first grade were used to create this composite. The weighted factor for each of the word identification and decoding measures was .167 and .333 for the comprehension measure. Nonresponders were those whose scores were below the 16th percentile on this reading composite (n = 23; 18.4%).

The second method of group classification was growth on the word identification fluency (WIF) progress monitoring data. At the end of the 14 weeks of Tier 2 tutoring, the locally normed WIF cut-points for future RD (derived from the Year 1 cohort) were used to identify nonresponders to Tier 2 instruction (intercept below = 21.41 words; slope below 16th percentile = 0.67 words/week). This procedure yielded 31 nonresponders (24.8%) to Tier 2 intervention.

Measures

Four academic measures were used in the current study that assessed word identification and decoding skills. In addition, there were nine cognitive measures that measured skills related to rapid naming, phonemic awareness, language, nonverbal problem-solving, and attention. For the purposes of this study, we consider cognitive skills that focus on the students' processing and do not involve reading. This is consistent with how others have operationalized cognitive skills in similar investigations (Fletcher et al., 2011). All measures were administered in the fall of first grade prior to random assignment to treatment.

Word identification. The *Word Identification* subtest from the *Woodcock Reading Mastery Test–Revised/Normative Update* (WRMT-R/NU; Woodcock, 1998), a norm-referenced test, asks children to read single words in list form. It consists of 100 words ordered in difficulty. Split-half reliability exceeded .90 for the normative first-grade sample. The *Sight Word Reading Efficiency* scale from the *Test of Word Reading Efficiency* (TOWRE; Torgesen, Wagner, & Rashotte, 1997) is a norm-referenced measure of sight word reading accuracy and fluency, assessing the number of real words accurately read in 45 s. It consists of 104 words ordered in difficulty. Split-half reliability exceeded .91 for the normative first-grade sample.

Decoding. The *Word Attack* subtest from the WRMT-R/NU (Woodcock, 1998) evaluates children's ability to pronounce pseudowords presented in list form. It contains 45 nonsense words, ordered from most easy to most difficult. Split-half reliability exceeded .90 for the normative first-grade sample. The *Phonemic Decoding Efficiency* subtest from the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1997) measures decoding accuracy and fluency, measuring the number of nonsense words accurately decoded in 45 s. It consists of 63 words ordered in difficulty. Split-half reliability exceeded .90 for the normative first-grade sample.

Rapid naming. The *Rapid Letter Naming* subtest from the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner, Torgesen, & Rashotte, 1999) requires children to name six letters arranged in random order on two pages. Each page displays four rows with nine letters per row. The tester asks the child to name each letter, corrects any errors, and then asks the child to name the letters again. If the child cannot name all of them or makes more than 4 errors during testing, the examiner discontinues testing. The child's score is the number of seconds required to name the 36 letters. Test–retest reliability is .97 for 5- to 7-year olds. The materials, administration protocol, and scoring for the *Rapid Digit Naming* subtest from the CTOPP mirror those of the *Rapid Letter Naming* subtest except that single digits were presented rather than single letters. Test–retest reliability is .91 for 5- to 7-year olds.

Phonemic awareness. The CTOPP also includes a subtest on *Sound Matching* that requires children to match first and last sounds in words. For first sound matching, children are presented with a word and then are asked which of three words (depicted as pictures) start with the same sound. A similar procedure explores last sound matching. After three practice items, the test comprises 20 items. Test–retest reliability is .83.

Language. The *Listening Comprehension* subtest of the *Woodcock Diagnostic Reading Battery* (WDRB; Woodcock, 1998) assesses the ability to listen to a sentence and orally provide the missing word. For example, the test giver reads the sentence "*A dog barks, a cat* _____," and the test taker is asked to provide the missing word. Two practice items are given before increasingly difficult items are administered. Items are scored correct and incorrect, and the final scores are the number of correct responses. The test is terminated when six consecutive incorrect answers are given. The test developers reported reliability of .80 for students aged 5-18 years.

The *Oral Vocabulary* subtest from the *Woodcock–Johnson Psychoeducational Battery–Revised* (WJ-R; Woodcock, McGrew, & Mather, 2001) assesses the ability to provide synonyms and antonyms in response to stimulus words presented by the examiner. Split-half reliability exceeded .90 for a normative first-grade sample. The *Vocabulary* subtest on the *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999) assesses expressive vocabulary, verbal knowledge, and foundation of information with 42 items. The first four items present pictures; the student identifies the object in the picture. For remaining items, the tester says a word that the student defines. Responses are awarded a score 0, 1, or 2 depending on quality. Testing is discontinued after five consecutive scores of 0. The score is the total number of points. As reported by Zhu (1999), split-half reliability is .86 to .87 at ages 6 to 7; the correlation with the Wechsler Intelligence Scale for Children–III Full Scale IQ is .72.

Nonverbal problem-solving. Matrix Reasoning on the WASI (Wechsler, 1999) measures nonverbal reasoning with four types of tasks: pattern completion, classification, analogy, and serial reasoning. Examinees look at a matrix from which a section is missing and complete the matrix by saying the number of or pointing to one of five response options. Examinees earn points by identifying the correct missing piece of the matrix. Testing is discontinued after four errors on five consecutive items or after four consecutive errors. The score is the number of correct responses. As reported by the test developer, reliability is .94 for 8-year-olds; the correlation with the Wechsler Intelligence Scale for Children–III Full Scale IQ is .66.

Attentive behavior. The *Strengths and Weaknesses of ADHD Symptoms and Normal Behavior* (SWAN; Swanson et al., 2004) is an 18-item teacher rating scale (www.adhd.net). Items from the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychiatric Association [APA], 1994) criteria for attention-deficit/hyperactivity disorder are included for inattention (Items 1-9) and hyperactivity/impulsivity (Items 10-18). Items are rated on a scale of 1 to 7 (1 = *far below*, 2 = *below*, 3 = *slightly below*, 4 = *average*, 5 = *slightly above*, 6 = *above*, 7 = *far above*). In the present study, we report data for teacher and tutor ratings on the inattentive behavior subscale. Using the nine relevant items, we used a principal components factor analysis to create a weighted composite variable of attentive behavior, or the ability to maintain focus. The SWAN has been shown to correlate well with other dimensional assessments of behavior related to inattention. Coefficient alpha on the representative sample was .96.

Results

We sought to examine whether the profiles of nonresponders were significantly different from those who did response to Tier 2 intervention, as well as how these comparisons might differ based on the criteria used to classify nonresponders. Profiles of responders and nonresponders were compared using a repeated measures analysis of variance (ANOVA). Specifically, two separate analyses were run for each of the academic and cognitive profiles, based on groups being identified using the two sets of classification described (i.e., reading composite and WIF). For each analysis, the between-subjects factor was response group (responders vs. nonresponders), and the within-subjects variables were the academic or cognitive measures.

The main effect for response group, referred to as the *elevation* effect, represents differences between groups averaged across all measures. The main effect for measures, referred to as the *flatness* effect, represents differences among the measures averaged across the groups. Within profile analysis, the interaction between group and measures, referred to as the *shape* effect, represents differences in the shape of the profile across groups (see Tabachnick & Fidell, 2006). Means, standard deviations, and correlations for all variables are presented in Table 1. Based on the sample of 125 students, raw scores were transformed for each of the variables to z scores ($M = 0.00$, $SD = 1.00$).

Academic Profiles

To explore the relationships between the academic measures (word identification, word attack, sight word efficiency, phonemic decoding) and response groups, we conducted a profile analysis. Two repeated measure ANOVAs were performed to explore elevation, flatness, and shape effects on the four academic measures across the response groups (responders vs. nonresponders). The first analysis included the reading composite group classification as the between-subjects variable, with pretest z scores for the academic measures as the within-subjects variables, whereas the second analysis was conducted with the same within-subjects variables, but with WIF group classification as the between-subjects variable.

Significant *elevation* effects were found using the reading composite, $F(1, 123) = 107.122$, $p < .001$, $\eta_p^2 = .466$, and WIF criteria, $F(1, 123) = 67.121$, $p < .001$, $\eta_p^2 = .353$. These findings indicate that groups differed from one another on their academic profiles when scores were averaged across dimensions. The effect of *flatness* was found to be nonsignificant in the profile using the reading composite for classification, $F(3, 369) = .385$, $p = .730$, $\eta_p^2 = .003$, and the WIF criteria, $F(3, 369) = 1.257$, $p = .289$, $\eta_p^2 = .010$. That is to say, when performance was aggregated across groups, there were no observed differences across measures.

The test of *shape* was nonsignificant when using the reading composite criteria, $F(3, 369) = .964$, $p = .399$, $\eta_p^2 = .008$. The *shape* effect was significant when using the WIF criteria, $F(3, 369) = 4.949$, $p = .004$, $\eta_p^2 = .039$, indicating deviation from parallelism across groups on the measures. These shape effects indicate the possibility of unique strengths and weaknesses across measures as a function of group.

As evident in Figure 1, the responders group exhibited relative strength on all measures regardless of the group classification criteria being used, whereas relative weaknesses were present among the nonresponders group. Because the elevation effects were striking and because many studies have compared univariate differences between responder groups, we computed effect sizes for each variable to help evaluate the magnitude of the differences (see Table 2). This was done by dividing the difference between group means by the pooled standard deviation across the two groups being compared (responders and nonresponders; Hedges & Olkin, 1985). We were unable to compare univariate differences between the groups based on classification criteria (i.e., nonresponders reading composite vs. nonresponders WIF) due to the nonindependence of the samples.

Cognitive Profiles

The same procedure was used to explore the relationships between the cognitive measures (rapid letter naming, rapid

Table 1. Means, Standard Deviations, and Correlations on Sample (n = 125).

	Raw		Transformed range[a]		Correlations												
Subscale	*M*	*SD*	Minimum	Maximum	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Word identification[b]	19.06	8.01	−2.38	2.62	—												
2. Word attack[c]	5.98	4.43	−1.35	3.61	.64**	—											
3. Sight word efficiency[d]	15.18	5.81	−2.61	2.38	.83**	.54**	—										
4. Phonemic decoding[e]	4.44	3.76	−1.18	2.81	.61**	.59**	.59**	—									
5. Rapid letter naming[f]	−66.00	26.05	−4.34	2.53	.18*	.11	.23*	.17	—								
6. Rapid digit naming[g]	−68.25	24.20	−3.38	2.82	.21*	.17	.31**	.23*	.68**	—							
7. Sound matching[h]	11.80	4.97	−2.17	1.65	.59**	.48**	.60**	.40**	.15	.29*	—						
8. Listening comp[i]	13.32	4.28	−2.88	2.26	.23*	.20*	.19*	.23*	.04	.01	.21*	—					
9. Oral vocabulary[j]	7.85	3.90	−2.01	2.35	.45**	.43**	.38**	.40**	−.05	.02	.42**	.55**	—				
10. Vocabulary[k]	18.34	6.66	−2.30	2.65	.19*	.13	.19*	.23*	−.01	−.05	.08	.51**	.38**	—			
11. Nonverbal PS[l]	6.71	4.23	−1.35	3.38	.33**	.33**	.20*	.21*	.01	−.06	.25*	.31**	.33**	.22*	—		
12. Attention teacher[m]	−0.05	0.91	−2.34	2.21	.50**	.38**	.49**	.37**	.07	.17	.50**	.12	.35**	.13	.22*	—	
13. Attention tutor[n]	−3.07	0.73	−2.74	2.54	.38**	.28*	.39**	.33**	.12	.25*	.38**	.09	.29*	.18*	.07	.46**	—

[a]Raw scores were transformed to z scores (M = 0.00, SD = 1.00); PS = problem-solving. [b]Woodcock Diagnostic Reading Battery (WDRB). [c]Woodcock Diagnostic Reading Battery (WDRB). [d]Test of Word Reading Efficiency (TOWRE). [e]Test of Word Reading Efficiency (TOWRE). [f]Comprehensive Test of Phonological Processing (CTOPP). [g]Comprehensive Test of Phonological Processing (CTOPP). [h]Comprehensive Test of Phonological Processing (CTOPP). [i]Woodcock Diagnostic Reading Battery (WDRB). [j]Woodcock–Johnson III (WJ-III). [k]Wechsler Abbreviated Scale of Intelligence (WASI). [l]Matrix Reasoning—Wechsler Abbreviated Scale of Intelligence (WASI). [m]Attention subscale—Strengths and Weaknesses of ADHD Symptoms and Normal Behavior (SWAN). [n]Attention subscale—Strengths and Weaknesses of ADHD Symptoms and Normal Behavior (SWAN).
*p < .05. **p < .001.

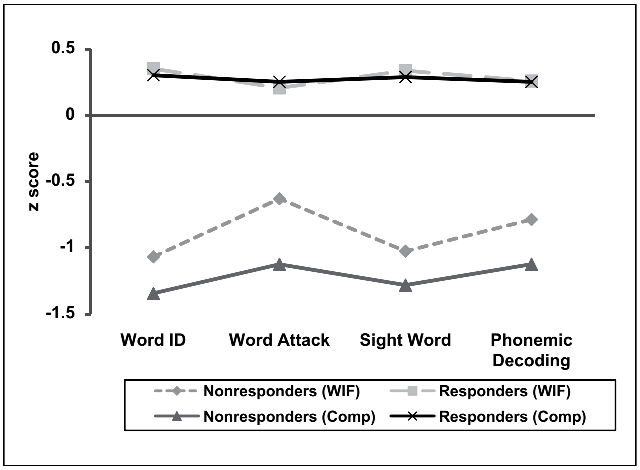

Figure 1. Academic profiles—z scores on four academic measures.
Note. WIF = word identification frequency; Comp = first-grade reading composite.

digit naming, sound matching, listening comprehension, oral vocabulary, vocabulary, nonverbal problem-solving, ratings of inattention by teacher and tutor) and response groups. Two repeated-measures ANOVAs were performed to explore elevation, flatness, and shape effects on the nine cognitive measures across the response groups (responders vs. nonresponders), separately examining the reading composite group classification and the WIF classification as the between-subjects variable.

Significant differences were found among groups when scores were averaged across all measures. *Elevation* effects were found using the reading composite, $F(1, 123) = 66.303$, $p < .001$, $\eta_p^2 = .350$, and WIF criteria, $F(1, 123) = 42.767$, $p < .001$, $\eta_p^2 = .258$. The effect of *flatness*, examining performance aggregated across group, was found to be nonsignificant in the profile using the reading composite for classification, $F(8, 984) = .416$, $p = .856$, $\eta_p^2 = .003$, and the WIF criteria, $F(8, 984) = .890$, $p = .496$, $\eta_p^2 = .007$.

Again, the test of *shape* was nonsignificant when using the reading composite criteria, $F(8, 984) = 1.041$, $p = .396$, $\eta_p^2 = .008$. However, the WIF criteria resulted in profiles that demonstrated *shape* effects, or deviated significantly from parallelism in performance across cognitive measures, $F(8, 984) = 3.505$, $p = .003$, $\eta_p^2 = .028$.

Similar to the academic profiles, Figure 2 presents cognitive profiles that show relative strength across all measures in the responder group using both sets of classification criteria. The nonresponders' group based on reading composite classification demonstrates relative weaknesses across all measures, whereas the nonreponders based on WIF appear to have a more diverse pattern of strengths and weaknesses. Effect sizes were computed for each variable to help evaluate the magnitude of the differences (see Table 2).

Table 2. Effect Sizes for Academic and Cognitive Variables as a Function of Response Group.

	Nonresponders vs. Responders	
Variable	Reading composite	WIF
Word identification[a]	−2.137	−1.795
Word attack[b]	−1.626	−0.895
Sight word efficiency[c]	−1.977	−1.685
Phonemic decoding[d]	−1.624	−1.169
Rapid letter naming[e]	−0.824	−0.559
Rapid digit naming[f]	−1.121	−0.799
Sound matching[g]	−1.253	−1.396
Listening comp[h]	−0.552	−0.200
Oral vocabulary[i]	−1.195	−0.439
Vocabulary[j]	−0.691	−0.420
Nonverbal problem-solving[k]	−0.765	−0.401
Attention teacher[l]	−0.926	−1.293
Attention tutor[m]	−0.677	−0.743

Note. WIF = word identification fluency.
[a]Woodcock Diagnostic Reading Battery (WDRB). [b]Woodcock Diagnostic Reading Battery (WDRB). [c]Test of Word Reading Efficiency (TOWRE). [d]Test of Word Reading Efficiency (TOWRE). [e]Comprehensive Test of Phonological Processing (CTOPP). [f]Comprehensive Test of Phonological Processing (CTOPP). [g]Comprehensive Test of Phonological Processing (CTOPP). [h]Woodcock Diagnostic Reading Battery (WDRB). [i]Woodcock–Johnson III (WJ-III). [j]Wechsler Abbreviated Scale of Intelligence (WASI). [k]Matrix Reasoning—Wechsler Abbreviated Scale of Intelligence (WASI). [l]Attention subscale—Strengths and Weaknesses of ADHD Symptoms and Normal Behavior (SWAN). [m]Attention subscale—Strengths and Weaknesses of ADHD Symptoms and Normal Behavior (SWAN).

Discussion

In the present study, we sought to explore the ways in which unresponsiveness is defined and interpreted. Specifically, we examined profiles of responders and nonresponders to Tier 2 intervention in the first grade—this included an array of academic and cognitive skills. As students were identified as nonresponders on the basis of low achievement, it was expected that we might find more variability across groups when examining their performance on measures of language and cognitive processing than on measures of academic achievement. We also conducted two separate profile analyses to examine differences in unresponsiveness based on the criteria used to classify students as adequate and inadequate responders (i.e., end of first grade reading composite vs. progress monitoring WIF data).

Classification of Unresponsiveness

Inconsistent methods for defining unresponsiveness can result in having quite different groups of students who move to more intensive intervention within the RTI framework (Tier 3), as well as influence decisions about who is identified as having a reading disability. The variability in

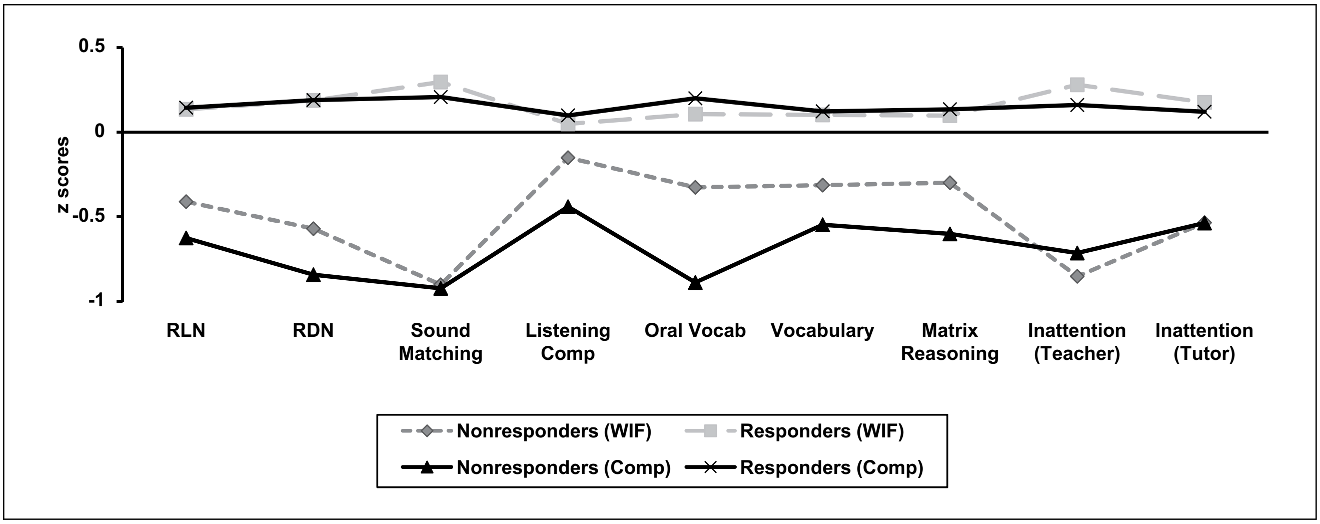

Figure 2. Cognitive profiles—z scores on nine cognitive measures.
Note. WIF = word identification frequency; Comp = first-grade reading composite.

classification criteria that has been reported in past research has resulted in little agreement about what constitutes adequate response to intervention (L. S. Fuchs et al., 2004; Good et al., 2001; McMaster et al., 2005; O'Connor & Klingner, 2010; Torgesen et al., 2001). Are these RTI identification methods resulting in similar decisions about unresponsiveness? When unresponsiveness is used for the purposes of identifying reading disabilities and assigning delivery of intensive intervention services (Tier 3), there is the need to understand *who* we consider to be nonresponders. The results from the present study demonstrate that the criteria used to classify adequate and inadequate responders to intervention (reading composite vs. WIF) result in different profiles for these groups. Using an end-of-first-grade reading composite, 18.4% of the students in the sample were classified as nonresponders, whereas 24.8% of students were classified when using local norms of WIF.

The use of WIF progress monitoring data appears to provide a more sensitive measurement of students' response to the intervention they are receiving at Tier 2 (Compton et al., 2010; Compton et al., 2012). Compton and colleagues (2012) found that the addition of Tier 2 progress monitoring WIF data, as were used in the current study, increased specificity (i.e., true negative rate) although it did not improve the predictive accuracy of the model. This is consistent with current findings that show that more students are identified as inadequate responders when the classification criteria use local norms, as compared with a composite of nationally normed measures. As such, it may be that WIF criteria lead to an overly inclusive identification method.

Profiles of Adequate and Inadequate Responders

Profile analyses revealed significant elevation effects for academic and cognitive profiles, regardless of the classification criteria being used. Responders' profiles differ from those of nonresponders when scores are averaged across domains, which is consistent with what we know about the characteristics of students who do not respond adequately to Tier 2 interventions (Al Otaiba & Fuchs, 2002; 2006; Fletcher et al., 2011; Nelson et al., 2003). Interestingly, only the analyses comparing groups using the WIF classification revealed significant shape effects for academic and cognitive profiles.

It appears that the primary difference between the two nonresponder groups is that the group identified using WIF criteria contains students who have some relative strengths in decoding, listening comprehension, oral vocabulary, vocabulary, and nonverbal reasoning, the majority of which are not skills assessed by WIF progress monitoring. Although nonresponders presented these strengths when classified using WIF criteria, they appear to perform consistently (poorly) across domains when assessed using a reading composite at the end of first grade.

One possible explanation for the variability found within the group of nonresponders identified using the WIF criteria is simply that we are selecting a larger number of students (24.8%, as compared with 18.4% with the reading composite criteria). Although using WIF as an RTI identification method may increase the chance that we are picking up all nonresponders, we are also increasing the likelihood of identifying students with higher levels of performance. Because using a reading composite as an RTI identification method identifies a smaller number of students, it is possible that these students are weaker across academic and cognitive skills, thus representing the lower end of the distribution.

Indeed, if we examine the overlap of students identified as nonresponders using both sets of classification criteria in our sample of 125 first graders, 28.8% (n = 31) were identified as unresponsive using either one of the identification methods. When we examine the students identified as nonresponders using WIF criteria, 42% no longer met criteria for unresponsiveness when using the end of first grade reading composite. Specifically, of the total sample, 71.2% (n = 89) were adequate responders, 14.4% (n = 18) met criteria for unresponsiveness for WIF and the reading composite, whereas 10.4% (n = 13) met WIF criteria only, and 4% (n = 5) met composite criteria only. We conducted univariate analyses to examine potential differences between students who were represented in only one group (WIF only vs. composite only). Students who met the criteria for unresponsiveness for the reading composite criteria *only* performed significantly worse than those who met the WIF criteria only on measures on decoding, oral vocabulary, listening comprehension, and nonverbal problem-solving. It may be that students with weaknesses in these areas had stronger performance on word reading measures (i.e., WIF progress monitoring) as compared with the reading comprehension component of the assessments administered at the end of first grade (i.e., reading composite). That is to say, these findings may point to potentially important considerations related to identification and placement of students in appropriately targeted interventions.

Limitations and Implications

Although these findings contribute to our understanding of unresponsiveness, this study is not without limitation. We have a relatively small sample size that did not allow for more advanced modeling of the profiles of nonresponders. Furthermore, the field has not yet come to consensus on criteria for determining unresponsiveness and, as such, we selected two commonly used methods of classifying adequate versus inadequate response. The focus of our Tier 2 intervention was reading at the word-level, and this certainly does not address the full range of reasons why a student may struggle with reading or fail to respond to instruction. And finally, as previously mentioned, statistical comparisons across groups (i.e., nonresponders based on reading composite vs. WIF data) were not possible due to the interdependence of the samples. Thus, we cannot make claims related to one group demonstrating clear strengths or weaknesses on specific academic or cognitive skills as compared with the other.

Overall, the current findings indicate that the criteria being using to identify nonresponders may suggest a slightly different set of skills with which the students struggle. Further examination of these differences is necessary to effectively target skills within our interventions. If we are to move students into Tier 3 interventions, we must be providing the appropriate level of individualization to meet their needs (Compton et al., 2012; Denton, Fletcher, Anthony, & Francis, 2006; Wanzek & Vaughn, 2009; Young, 2008). The current findings suggest that we may need to consider multiple sources of data to have a complete understanding of unresponsiveness. From the perspective of learning disability identification, students who move on to Tier 3 reading intervention are most likely students with RD, or those who will go on to be identified with an RD. With the understanding that the group of students we deem to be unresponsive differs based on classification criteria, there are certainly important future considerations for how we identify learning disabilities within RTI models.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by Grant R324G060036 from the Institute of Education Sciences, U.S. Department of Education, and Core Grant HD15052 from the National Institute of Child Health and Human Development to Vanderbilt University. Statements do not reflect the position or policy of these agencies, and no official endorsement by them should be inferred.

References

Ackerman, P. T., Dykman, R. A., & Gardner, M. Y. (1990). ADD students with and without dyslexia differ in sensitivity to rhyme and alliteration. *Journal of Learning Disabilities*, *23*, 279–283.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities*, *39*, 414–431.

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education*, *23*, 300–316.

Biemiller, A. (1977). Relationships between oral reading rates for letters, words, and simple text in the development of reading achievement. *Reading Research Quarterly*, *13*, 223–253.

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders, and reading retardation. *Journal of Child Psychology and Psychiatry*, *31*, 1027–1050.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., . . .Crouch, R. C. (2010). Selection at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, *102*, 327–340.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*, *98*, 394–409.

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., . . .Bouton, B. D. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities*, *45*, 204–216.

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

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. New York, NY: Guilford Press.

Fletcher, J. M., Stuebing, K. K., Barth, A. E., Denton, C. A., Cirino, P. T., Francis, D. J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to reading intervention. *School Psychology Review*, *40*, 3–22.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves study. *Journal of Educational Psychology*, *88*, 3–17.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, N. G. (2008). Making "secondary intervention" work in a three-tier Responsiveness-to-Intervention model: Findings from the first-grade longitudinal reading study of the National Research Center on Learning Disabilities. *Reading and Writing Quarterly: An Interdisciplinary Journal*, *21*, 413–436.

Fuchs, D., & Deshler, D. K. (2004). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice*, *20*, 129–136.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multi-level prevention. *Exceptional Children*, *78*(3), 263–279.

Fuchs, D., Fuchs, L. S., & Vaughn, S. (2008). *Response to intervention: A framework for reading educators*. Newark, DE: International Reading Association.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children*, *71*, 7–21.

Fuchs, L. S., Fuchs, D., Craddock, C., Hollenbeck, K. N., Hamlett, C. L., & Schatschneider, C. (2008). Effects of small-group tutoring with and without validated classroom instruction on at-risk students' math problem solving: Are two tiers of prevention better than one? *Journal of Educational Psychology*, *100*, 491–509.

Gilbert, J. K., Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Barquero, L. A., & Cho, E. (2013). Efficacy of a first-grade responsiveness-to-intervention prevention model for struggling readers. *Reading Research Quarterly*, *48*, 135–154.

Good, R. H., Simmons, D., Kame'enui, E., & Chard, D. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, *5*, 257–288.

Hall, S., & Moats, L. C. (1999). *Straight talk about reading: How parents can make a difference during the early years*. Lincolnwood, IL: NTC/Contemporary Publishing Group.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Lovett, M. W. (1984). A developmental perspective on reading dysfunction: Accuracy and rate criteria in the subtyping of dyslexic children. *Brain & Language*, *22*, 67–91.

Lyon, G. R. (1995). Research initiatives in learning disabilities: Contributions from scientists supported by the National Institute of Child Health and Human Development. *Journal of Child Neurology*, *10*, 120–126.

Mathes, P. G., Howard, J. K., Allen, S. H., & Fuchs, D. (1998). Peer-assisted learning strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly*, *33*, 62–94.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, *71*, 445–463.

Menyuk, P., & Chesnick, M. (1997). Metalinguistic skills, oral language knowledge, and reading. *Topics in Language Disorders*, *17*, 75–87.

National Institute of Child Health and Human Development. (2000). *Report of the national reading panel: Teaching students to read: An evidenced-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00–4754.). Washington, DC: U.S. Government Printing Office.

National Reading Panel. (2000). *National reading panel report: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (No. NIH Pub. 00-4769). Washington, DC: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development.

Nelson, R. J., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice*, *18*, 255–267.

O'Connor, R. E., & Klingner, J. (2010). Poor responders in RTI. *Theory Into Practice*, *49*, 297–304.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academic Press.

Speece, D. L., & Case, L. P. (2001). Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology*, *93*, 735–749.

Swanson, J., Schuck, S., Mann, M., Carlson, C., Hartman, K., Sergeant, J., . . .McCleary, R. (2004). *Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and the SWAN rating scales*. Available from www.adhd.net

Tallal, P. (1990). Fine-grained discrimination deficits in language-learning impaired children are specific neither to auditory modality nor to speech perception. *Journal of Speech and Hearing Research*, *33*, 616–617.

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.

Torgesen, J. K. (2000). Individual responses in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, *15*, 55–64.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. S., & Conway, T. (2001). Intensive

remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, *34*, 33–58.
Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, *27*, 276–286.
Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). *Test of word reading efficiency*. Austin, TX: Pro-Ed.
Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to treatment: Rationale and assumptions. *Learning Disabilities Research & Practice*, *18*, 13–146.
Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children*, *69*, 391–409.
Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: Pro-Ed.
Wanzek, J., & Vaughn, S. (2009). Students demonstrating persistent low response to reading intervention: Three case studies. *Learning Disabilities Research & Practice*, *24*, 151–163.
Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corp.
Wolf, M. (1991). Naming speed and reading: The contribution of the cognitive neurosciences. *Reading Research Quarterly*, *26*, 123–141.
Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *3*, 415–438.
Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests–Revised/Normative Update*. Circle Pines, MN: American Guidance Service.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III tests of psychoeducational ability*. Itasca, IL: Riverside.
Young, C. L. (2008). *The effects of group size on reading outcomes of identified nonresponders* (Unpublished doctoral dissertation). Vanderbilt University, Nashville, TN.
Zhu, J. (1999). *WASI manual*. San Antonio, TX: The Psychological Corporation.