Published in final edited form as:

School Psych Rev. 2014 December ; 43(4): 407–427. doi:10.17105/SPR-13-0052.1.

Cognitive Attributes of Adequate and Inadequate Responders to Reading Intervention in Middle School

Jeremy Miciak, PhD,

An assistant research professor at the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston. He holds a doctorate in special education from The University of Texas at Austin. His research interests include issues of definition, identification, and treatment of learning disabilities, particularly among students from diverse cultural and linguistic backgrounds

Karla K. Stuebing, PhD,

Retired research professor at the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston. Her research interests include measurement and classification issues in the fields of reading and learning disabilities

Sharon Vaughn, PhD,

H. E. Hartfelder/Southland Corp. Regents Chair in Human Development and the Executive Director of The Meadows Center for Preventing Educational Risk, an organized research unit at The University of Texas at Austin. She is the recipient of numerous awards including the Council of Exceptional Children (CEC) Research Award, the American Educational Research Association Special Interest Group Distinguished Researcher Award, The University of Texas Distinguished Faculty Award and Outstanding Researcher Award, and the Jeannette E. Fleischner Award for Outstanding Contributions in the Field of LD from CEC. She is the author of more than 35 books and 250 research articles

Greg Roberts, PhD,

Director of the Vaughn Gross Center for Reading and Associate Director of The Meadows Center for Preventing Educational Risk. He directs all data related activities for the Centers. He has been principal investigator or coinvestigator on over a dozen research projects funded by the Institute of Education Sciences, National Institutes of Health, and others. Trained as an educational research psychologist, with expertise in quantitative methods, Dr. Roberts has directed evaluation projects of programs in education, social services, and health care. He has over 50 publications in peer-reviewed journals. He has published in multidisciplinary Tier 1 journals using structural equation models, meta-analysis, and multilevel models

Amy Elizabeth Barth, PhD, and

An assistant professor in special education at the University of Missouri-Columbia. Dr. Barth, a speech-language pathologist, is currently completing the third year of a 5-year K08 Mentored Clinical Scientist Research Career Development Award from the National Institute of Child Health and Human Development. Her interests include the identification and treatment of students with language and reading disabilities

Jack M. Fletcher, PhD

Hugh Roy and Lillie Cranz Cullen Distinguished Professor and Chair, Department of Psychology, at the University of Houston. Dr. Fletcher, a child neuropsychologist, has conducted research on children with learning and attention disorders, as well as brain injury. He served on the 2002 President's Commission on Excellence in Special Education. Dr. Fletcher received the Samuel T. Orton Award from the International Dyslexia Association in 2003 and was a corecipient of the Albert J. Harris Award from the International Reading Association in 2006

Abstract

No studies have investigated the cognitive attributes of middle school students who are adequate and inadequate responders to Tier 2 reading intervention. We compared students in Grades 6 and 7 representing groups of adequate responders ($n = 77$) and inadequate responders who fell below criteria in (a) comprehension ($n = 54$); (b) fluency ($n = 45$); and (c) decoding, fluency, and comprehension (DFC; $n = 45$). These students received measures of phonological awareness, listening comprehension, rapid naming, processing speed, verbal knowledge, and nonverbal reasoning. Multivariate comparisons showed a significant Group-by-Task interaction: the comprehension-impaired group demonstrated primary difficulties with verbal knowledge and listening comprehension, the DFC group with phonological awareness, and the fluency-impaired group with phonological awareness and rapid naming. A series of regression models investigating whether responder status explained unique variation in cognitive skills yielded largely null results consistent with a continuum of severity associated with level of reading impairment, with no evidence for qualitative differences in the cognitive attributes of adequate and inadequate responders.

Previous evaluations of the cognitive profiles of struggling readers have primarily focused on young children struggling to acquire foundational reading skills such as phonological awareness, basic decoding skills, and reading fluency (Fletcher et al., 2011; McMaster, Fuchs, Fuchs, & Compton, 2005; Stage, Abbott, Jenkins, & Beminger, 2003). However, as students grow older and are confronted with more complex and cognitively demanding texts, specific difficulties in reading comprehension may emerge in students with adequate decoding and fluency skills, marked primarily by limitations in listening comprehension and vocabulary (Catts, Hogan, & Adlof, 2005). Therefore, evaluations of the cognitive processes of younger struggling readers may not generalize to older struggling readers, among whom comprehension difficulties may be more prominent. In this study, we investigated the cognitive attributes of middle school students who showed adequate and inadequate responses to a Tier 2 reading intervention, including adolescents with specific difficulties with reading comprehension.

COGNITIVE ATTRIBUTES OF ADOLESCENT AND LATE ELEMENTARY SCHOOL STRUGGLING READERS

Considerable research has investigated the cognitive skills underlying adolescent literacy, particularly for struggling readers. For example, Catts, Adlof, and Weismar (2006) investigated the language comprehension and phonological awareness skills of eighth-grade

students with specific achievement deficits in reading comprehension or basic decoding. Poor comprehenders showed relative deficits in receptive vocabulary and grammatical understanding. In contrast, poor decoders showed relative deficits on measures of phonological awareness. Compton, Fuchs, Fuchs, Lambert, and Hamlett (2012) investigated the cognitive profiles of late elementary school students identified as having learning disabilities (LD) with specific deficits in reading comprehension or word reading or with mathematics deficits. Specific to reading, Compton et al. (2012) found that students with deficits in reading comprehension showed pronounced, specific deficits in language, corroborating the findings of Catts et al. (2006). In contrast, students with word reading deficits showed relative deficits in language and working memory, a domain not assessed by Catts et al. (2006).

Lesaux and Kieffer (2010) studied the language and reading skills of adolescents with comprehension deficits, in possible combination with other reading deficits. They identified three unique skill profiles for poor comprehenders using latent class analysis: slow word callers, automatic word callers, and globally impaired readers. Slow word callers showed above average decoding skills but impaired fluency; automatic word callers had above average decoding skills with adequate fluency. Globally impaired readers showed deficits in all areas. Despite differences in decoding and fluency, all three poor comprehender groups showed deficits in vocabulary, replicating the findings of Catts et al. (2006) and Compton et al. (2012) linking language and reading comprehension.

Barth, Catts, and Anthony (2009) investigated the reading and cognitive abilities underlying fluency, which is a third domain of reading. Confirmatory factor analysis and structural equation modeling indicated that word and text reading fluency constituted a single latent factor, a finding consistent with subsequent studies investigating component reading skills among middle school students (Cirino et al., 2012). After the authors controlled for differences in nonverbal intelligence, individual differences in decoding, language comprehension, and rapid naming explained over 80% of the variance in reading fluency performance (Barth et al., 2009). Of these three factors, rapid naming was most related to reading fluency, uniquely explaining 25% of the variance in reading fluency.

COGNITIVE CHARACTERISTICS OF INADEQUATE RESPONDERS

Studies of the cognitive attributes of struggling adolescent readers have identified a number of potential correlates of specific reading deficits, including language and vocabulary, rapid naming, and phonological awareness. However, one limitation to these descriptive studies is that none evaluated the cognitive characteristics of adolescents who did not respond to intervention. Academic underachievement has many potential causes, including limited academic opportunity. Response to intervention (RTI) service delivery models ensure that all students have been provided appropriate opportunities to learn through the systematic implementation of a generally effective intervention (Gresham et al., 2005; VanDerHeyden & Bums, 2010). Thus, inadequate responders represent an interesting subgroup of students showing intractability in learning academic skills (Fletcher & Vaughn, 2009).

Previous studies investigating correlates of inadequate responders to intervention were limited to students in elementary school grades. Al Otaiba and Fuchs (2002) reviewed 23 studies that included sufficient descriptive data on inadequate responders to permit analysis. Across all studies, seven child characteristics were significantly associated with inadequate responder status: phonological awareness, phonological memory, rapid naming, intelligence, attention or behavior, orthographic processing, and demographic characteristics. Phonological awareness was most frequently investigated and showed the closest association with inadequate responder status. Intelligence was less frequently associated with responder status.

Nelson, Benner, and Gonzalez (2003) extended the work of Al Otaiba and Fuchs (2002) using meta-analytic procedures. Thirty studies reported sufficient quantitative information to estimate at least one effect size between learner characteristics and responder status. The results indicated that rapid naming (weighted mean $Z_r = 0.51$), problem behavior (weighted mean $Z_r = 0.46$), phonological awareness (weighted mean $Z_r = 0.42$), alphabetic principle (weighted mean $Z_r = 0.35$), memory (weighted mean $Z_r = 0.31$), and IQ (weighted mean $Z_r = 0.26$) were significantly associated with treatment responder status. In contrast to the findings of Al Otaiba and Fuchs (2002), demographic characteristics were not significantly associated with responder status.

COGNITIVE ASSESSMENT AFTER INADEQUATE RTI

Partly because of the relation of cognitive processes and reading skills, comprehensive academic and cognitive testing after a determination of inadequate RTI has been recommended as an alternative to the longstanding practice of evaluating cognitive skills in children with suspected LD (Fiorello, Flale, & Snyder, 2010; Haleet al., 2010). This evaluation would be used for LD identification and subsequent treatment planning. However, the utility of such an evaluation has been questioned (Fletcher et al., 2011). At issue is whether potential differences discovered through cognitive assessment contribute unique diagnostic or prescriptive information not attainable through achievement tests alone.

Studies investigating correlates of inadequate RTI in early elementary school have found that adequate and inadequate responders can be differentiated on a wide range of skills, including initial reading skill, phonological and orthographic awareness, rapid naming, vocabulary, and oral language (Fletcher et al., 2011; Stage et al., 2003; Vellutino, Scanlon, & Jaccard, 2003; Vellutino, Scanlon, Small, & Fanuele, 2006). However, the fundamental question is whether cognitive differences between adequate and inadequate responders reflect unique cognitive attributes associated with inadequate responder status or whether cognitive differences parallel differences in the severity of reading impairment. Recent studies have found a strong linear relation between the cognitive skills underlying reading and reading achievement, with a stepwise progression among groups reflecting the severity of reading impairment (Fletcher et al., 2011; Vellutino et al., 2006). If cognitive abilities lie on a continuum of severity that reflects the severity of reading impairment, this would raise questions about the utility of cognitive assessment. This is because cognitive differences would reflect the severity of academic impairment and relative standing on cognitive

measures would contribute no additional, meaningful data for identification or treatment purposes.

PURPOSE

The goal of this study was to investigate the cognitive attributes of adequate and inadequate responders to intervention in middle school. Recent studies of inadequate responders have used only word level and fluency measures to determine intervention responder status. The inclusion of a criterion measure of comprehension in this study allowed a comparison of the cognitive attributes of inadequate responders with specific deficits in reading comprehension versus inadequate responders with deficits in fluency and decoding.

Three research questions were addressed:

1. What cognitive attributes differentiate inadequate and adequate responders to supplemental reading intervention?
2. To what extent do the cognitive attributes of inadequate responders differ according to the assessed reading domain?
3. How well does responder status predict differences in cognitive attributes beyond those reflected by the severity of reading impairment?

We hypothesized that the results will be consistent with a continuum-of-severity hypothesis. Thus, inadequate and adequate responders will be differentiated across cognitive attributes; however, inadequate responders with similar reading deficits will not differ in cognitive performance. We further hypothesized that when the severity of reading impairment is accounted for, there will be no unique variance associated with responder status.

METHODS

This study was conducted with the approval of the institutional review boards of the respective universities. Participants were drawn from seven middle schools in two large urban cities in the Southwestern United States. Four of the schools were in two small districts in one city. These schools ranged in size from 633 to 1,300 students and drew students from a mix of urban and rural neighborhoods. The remaining three schools were from a large urban district. The school sizes ranged from 500 to 1,400 students. The schools reflected the demographic characteristics of urban centers in Texas. Four of the seven schools included a large number of minority and economically disadvantaged students. In addition, a large percentage of students at each school qualified for free or reduced-price lunch (range = 40%–86%).

Participants

The study participants were drawn from a multiyear intervention study investigating the effects of Tier 1, 2, and 3 reading interventions in sixth through eighth grades (see Vaughn, Cirino et al., 2010; Vaughn et al., 2011, 2012; Vaughn, Wanzek et al., 2010, for reports on intervention effectiveness across all 3 years of intervention). Sixth- and seventh-grade students were selected for the Tier 2 reading intervention based on their Texas Assessment

of Knowledge and Skills (TAKS) reading score from the previous year (when students were in fifth grade and sixth grade, respectively). Students who scored below a scale score of 2,150 were randomly assigned in a 2:1 ratio to the Tier 2 intervention or a business-as-usual control group. In fifth and sixth grades, a TAKS scale score of 2,100 is considered passing. A scaled score of 2,150 indicates that the lower band of the 95% confidence interval included a failing score. The cut point was selected to approximate the 30th percentile on other normed–referenced assessments of reading (Vaughn et al., 2011). In addition, students who underwent an alternative state assessment because of the requirements of their special education program and had significant deficits in reading were included. Students with moderate or severe disabilities who did not participate in general education classes were excluded. Data were collected in fall and spring of Year 1, as well as fall of Year 2, before any potential Tier 3 intervention.

A total of 326 sixth and seventh graders began the intervention. Sixteen students did not complete the intervention and were unavailable for assessment in spring of Year 1 (remaining $n = 310$). An additional 70 students were lost in the summer between Year 1 and Year 2 (remaining $n = 240$) and were unavailable for assessment in fall of Year 2. Students lost to summer attrition did not differ from students who remained on a multivariate analysis of variance (MANOVA) comparing performance on the three criterion reading measures, $F(3, 306) = 0.63, p > .05, \eta^2 = 0.01$. In addition, we excluded three students who scored two standard deviations below the population mean score on the Verbal Knowledge and Matrix Reasoning subtests of the Kaufman Brief Intelligence Test–Second Edition (KBIT-2; Kaufman & Kaufman, 2004) because of the possibility of general intellectual deficiencies. The final sample consisted of 237 students who completed the Tier 2 reading intervention in sixth grade ($n = 169$) or seventh grade ($n = 71$). The sample included a large number of students who received free or reduced-price lunch, as well as a large number of minority students. Participant demographic characteristics are presented in Table 1 and discussed in subsequent sections.

Intervention

All participants in this study were randomly assigned to the Tier 2 intervention and attended a supplementary reading intervention for one period (45–50 min) each day as part of their regular schedule for the entire school year (see Vaughn, Cirino et al., 2010; Vaughn, Wanzek et al., 2010, for a complete description). The intervention was conducted in groups of 10–15 students. Each interventionist participated in approximately 60 hours of professional development provided by the research team before delivering instruction. An additional 9 hours of professional development was provided throughout the year. Biweekly staff meetings and ongoing on-site feedback helped ensure high fidelity of implementation.

The standardized, multicomponent intervention addressed (a) word study, (b) reading fluency, (c) vocabulary, and (d) comprehension. The intervention lessons proceeded in three phases with different emphases. Phase 1 lasted 7–8 weeks and emphasized word study and fluency. Students engaged in structured partner reading and word study lessons from Archer, Gleason, and Vachon (2005a), in addition to daily vocabulary and comprehension instruction. Phase 2 lasted 17–18 weeks. During this phase, instruction emphasized

vocabulary and comprehension while fluency and word study skills from Phase 1 continued to be practiced. Similar to Phase 1 vocabulary instruction, students were introduced to new words and provided basic definitions, examples, and nonexamples. Students also studied parts of speech, as well as word relatives. Reading materials were drawn from social studies lessons provided by Archer, Gleason, and Vachon (2005b) and reading level-appropriate novels. Comprehension instruction focused on improving comprehension strategies, particularly question generation. As students read text, teachers provided explicit instruction on formulating literal questions, questions requiring information synthesis, and questions dependent on the application of background knowledge. Students also received instruction on identifying the main idea and summarizing text. Phase 3 comprised the final 8–10 weeks of the intervention. Word study and vocabulary mirrored Phase 2, whereas fluency and comprehension activities were embedded in novel study units developed by the research team.

Intervention teachers participated in approximately 60 hours of training and ongoing coaching. Fidelity of implementation was collected across the year for each interventionist by a trained member of the research team. Interventionists were observed up to five times, with a median of four observations per teacher. Interventionists were given scores on a three-point Likert-type scale, from 1 (*low*) to 3 (*high*), to rate the implementation of each intervention component, as well as the quality of implementation. The scale score represented a holistic assessment of the extent to which the teacher completed the required elements and procedures to meet the objectives of each component. Mean teacher scores ranged from 2.00–3.00 for fidelity of implementation and from 2.00–2.96 for quality of instruction, indicating good fidelity of implementation and quality of instruction, respectively.

After 1 year of intervention, Tier 2 treatment students in sixth grade showed gains (median $d = 0.16$) in basic reading, reading fluency, and reading comprehension in comparison with a business-as-usual control group. In contrast, students in seventh grade showed few statistically significant gains in comparison with a business-as-usual control group. The effect sizes found in this study compare favorably with effect sizes on standardized reading measures in other randomized controlled trials conducted in adolescents (Scammacca, Roberts, Vaughn, & Stuebing, in press). That the aggregate effects for middle school interventions are generally smaller than the effect sizes of similarly designed interventions in early elementary school is not surprising or indicative of generally ineffective interventions. Effect sizes on standardized measures for elementary and middle school students are not equivalent. Effect sizes for a full year of instruction on standardized reading measures drop from 0.97 between Grades 1 and 2 to 0.23 between Grades 6 and 7 (Lipsey et al., 2012). Furthermore, large aggregate effects are not essential to the present analysis because we were able to identify a large group of participants who showed adequate response ($n = 77$) and a large group who did not ($n = 160$). This allowed comparisons between groups and permitted investigation of inadequate responders, a population of interest precisely because of the intractability of their academic difficulties.

Criteria for Inadequate Responder Group Formation

Following Tier 2 intervention, we applied criteria for the identification of inadequate responders in three reading domains: decoding, reading fluency, and reading comprehension. The use of multiple criteria allows a comparison of the cognitive attributes of inadequate responders who did not meet criteria in different reading domains and may provide greater sensitivity than the application of a single criterion measure (Fletcher et al., 2011). In addition, assessment with psychometrically sound, standardized measures across reading domains allows for the identification of students who show deficits in a specific reading domain, which may not be possible if a determination of adequate response is based on curriculum-based measures only.

Inadequate responder status was defined as a posttest standard score below 91 (25th percentile) on the (a) Woodcock–Johnson III (WJ-III; Woodcock, McGrew, & Mather, 2001) basic reading composite; (b) Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999); or (c) WJ-III Passage Comprehension subtest. The cut point for the three norm-referenced measures was selected to align with previous studies investigating RTI (Fletcher et al., 2011; Vellutino et al., 2003, 2006). The use of multiple indicators may result in greater sensitivity and minimize false negatives. This is important because (a) single indicators of responder status show poor to moderate agreement in classification decisions (Barth et al., 2008; Case, Speece, & Molloy, 2003) and (b) false negatives are comparatively deleterious because students who may need further intervention will not be identified. Although many RTI models use slope or dual-discrepancy criteria for determinations of responder status (Fuchs & Deshler, 2007), there is little evidence that slope explains significant variance beyond final status for the identification of responder status, especially when considering a restricted range of reading ability, such as students screened into Tier 2 intervention (Schatschneider, Wagner, & Crawford, 2008; Tolar, Barth, Fletcher, Francis, & Vaughn, 2014). In addition, final status indicators directly answer the fundamental question confronting educators after Tier 2 intervention: Does this student require additional reading intervention?

The application of response criteria yielded 77 adequate responders (i.e., scored above criteria on all three measures) and seven subgroups of inadequate responders ($n = 160$), reflecting students identified through all possible combinations of the three criteria. Mean scores on criterion measures of reading are presented for all seven inadequate responder groups in Table 2. The largest subgroup of inadequate responders fell below the cut point in comprehension only (comprehension group; $n = 54$). A second large group fell below the cut point on decoding, fluency, and comprehension (DFC group; $n = 45$). A third, smaller group fell below criteria on fluency only (fluency group; $n = 19$). Eight students fell below the cut point in decoding only, whereas 34 students fell below cut points in two of the three criterion measures.

Measures and Procedures

The data presented in this article were collected at three times: fall and spring of Year 1, as well as fall of Year 2, prior to any potential Tier III treatment that the student may have received. Verbal knowledge was measured in September of Year 1. Academic performance

and nonverbal reasoning were assessed in May of Year 1, as part of the posttest battery. Phonological processing, listening comprehension, and processing speed were assessed in September of Year 2, prior to the start of Year 2 intervention. To address discrepant testing dates for cognitive measures, we used age-based standard scores for all cognitive measures except the Underlining Test, for which normative scores were unavailable. It was necessary to administer the verbal knowledge and nonverbal reasoning assessments in Year 1 of the larger study to screen for students with intellectual deficits, who may have been ineligible to continue the study. All other cognitive processing assessments were administered at a single time point, following Tier 2 intervention but prior to any subsequent Tier 3 intervention.

Cognitive Processing Tests—We selected cognitive measures that assessed student performance across multiple domains empirically implicated as correlates of inadequate responder status to intervention in reading (Nelson et al., 2003) or of constructs often associated with reading disabilities. We also examined models of cognitive processing commonly used as part of an assessment of cognitive processing strengths and weaknesses in children based on the Cattell–Hom–Carroll (CHC) theory. We did not assess visual processing abilities because research suggests a tenuous relationship with reading (Evans, Floyd, McGrew, & Leforgee, 2001; McGrew, 1983). In the sections that follow, we describe each cognitive processing variable and discuss its theoretical and empirical relation to reading and to models of cognitive processes.

Comprehensive Test of Phonological Processing: The cognitive measures included the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) Blending Phonemes, Elision, and Rapid Automated Naming–Letters (RAN-L) subtests. These measures were selected to assess phonological awareness, an indicator of auditory processing in the CHC model, and rapid letter naming skills, a measure used as an indicator of the CHC longterm retrieval factor. Both constructs have been identified as major correlates of poor reading among adolescents (Barth et al., 2009; Catts et al., 2006). The CTOPP is a nationally normed, individually administered test of phonological awareness and phonological processing. We administered three subtests: Blending Words, Elision, and RAN-L. The Blending Words and Elision subtests were used to calculate a phonological awareness composite. For students aged 8–17 years, the test–retest reliability coefficient is 0.72 for the Blending Words subtest and 0.79 for the Elision subtest. The RAN-L subtest is a measure of fluency in naming letters. The test–retest reliability coefficient for the RAN-L subtest for students aged 8–17 years is 0.72. Confirmatory factor analysis supports the construct validity of the CTOPP, and the administered subtests indicate the latent constructs of phonological awareness and rapid naming (Wagner et al., 1999). The three subtests show moderate correlations with criterion measures of reading (r^2 range = 0.16–0.75; Wagner et al., 1999).

Underlining Test: The Underlining Test (Doehring, 1968) was used to assess processing speed for verbal and nonverbal information because deficits in this domain are frequently found in students identified with LD (Wolff, 1993) and as a measure of the processing speed factor in the CHC model. The Underlining Test is an individually administered measure of processing speed. During standard administration, a target stimulus is presented at the top of

a page. Below, there are lines including the target stimulus and distracters. The student is asked to underline only the target stimuli as quickly and accurately as possible for 30 or 60 s. We administered four tests with different target stimuli, but we only used scores on the first three subtests for subsequent analyses because one subtest serves as a control for motor speed. The raw score for the three included subtests is the total number of correct stimuli underlined minus the number of errors. Raw scores were converted to z scores ($M = 0$, $SD = 1$). The converted z scores for the three subtests were then averaged, providing a mean z score.

KBIT-2: The KBIT-2 Verbal Knowledge and Matrices subtests were administered to assess vocabulary, nonverbal reasoning, and perceptual skills, representing the comprehension-knowledge and fluid reasoning CHC factors. The KBIT-2 is a nationally normed, individually administered measure of verbal and nonverbal intelligence (Kaufman & Kaufman, 2004). The Verbal Knowledge subtest requires the student to match stimulus pictures with a word or phrase. The verbal knowledge score was prorated for the verbal domain, allowing computation of both verbal and nonverbal standard scores. The Matrices subtest requires the student to select which picture best goes with the stimulus picture, which includes both meaningful and abstract images. For the age range of this study, test–retest reliability ranges from 0.80–0.93. Both the Matrices and Verbal Knowledge subtests correlate with other measures of intelligence, with adjusted correlations ranging from 0.79–0.84 for the verbal composite to other verbal measures and 0.47–0.81 for the nonverbal composite to other perceptual reasoning measures.

Group Reading Assessment and Diagnostic Evaluation: As an additional oral language measure, we administered the Listening Comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001). Listening comprehension is sometimes used as an indicator of aptitude among students with LD (Stanovich, 1991) and has been implicated as a correlate of poor reading among adolescents (Catts et al., 2006; Lesaux & Kieffer, 2010). We chose to include the Listening Comprehension subtest as an additional measure of the comprehension–knowledge CHC factor because of strong empirical and theoretical links among lexical knowledge, language development, and reading comprehension. The GRADE is a nationally normed, group-administered test of reading and listening comprehension. The test–retest reliability coefficient for sixth grade is 0.94, and the alternate-form reliability coefficient is 0.88. Concurrent validity between the GRADE and the Iowa Test of Basic Skills for reading ability ranges between 0.69 and 0.83 (Williams, 2001).

Measures to Determine Intervention Responder Status—We selected three norm-referenced assessments of reading to serve as criterion measures for a determination of responder status. The three measures assess three reading domains: decoding, fluency, and comprehension. The use of three psychometrically sound measures of reading across multiple domains results in improved sensitivity and permits identification of students with specific reading deficits.

TOWRE: The TOWRE is a nationally normed, individually administered test of reading fluency (Torgesen et al., 1999). We administered the Sight Word Efficiency and Phonemic Decoding Efficiency subtests. We used the composite score. Test–retest reliability for TOWRE subtests and composite scores for students aged 10–18 years ranges from 0.83–0.92. Concurrent correlations with the Woodcock Reading Mastery Tests–Word Attack and –Word Identification subtests range from 0.86–0.94.

WJ-III Test of Achievement: The WJ-III is a nationally normed, individually administered test of academic achievement (Woodcock et al., 2001). We administered the Letter–Word Identification, Word Attack, and Passage Comprehension subtests. The Letter–Word Identification and Word Attack subtests represent the basic reading composite. Reliability coefficients for students aged 11–14 years range from 0.91–0.96 for the basic reading composite and from 0.80–0.86 for the Passage Comprehension subtest. The concurrent correlation of the basic reading composite with the Basic Reading subtest of the Wechsler Individual Achievement Test is 0.82. The concurrent correlation of the Reading Comprehension subtest with the Reading Comprehension subtest of the Wechsler Individual Achievement Test is 0.79.

Additional Academic Measures—We also used three reading measures to evaluate the comparability of inadequate responder groups. These measures were used to determine whether groups were sufficiently similar to be combined for subsequent analyses only.

GRADE Reading Comprehension: The GRADE is a nationally normed, group-administered test of reading and listening comprehension (Williams, 2001). We administered the Grade 6 Passage Comprehension subtest. The GRADE produces a stanine score for the Passage Comprehension subtest, but for purposes of this study, we prorated the raw score to derive a standard score for the GRADE comprehension composite.

AIMSweb Reading Maze: The AIM-Sweb Reading Maze is a group-administered, 3-min test of reading fluency and comprehension (Shinn & Shinn, 2002). During standard administration, the student is required to read a 150- to 400-word passage and select the correct omitted word from among three choices as quickly and accurately as possible. After the first sentence, every seventh word of text is omitted. Within our sample, the mean intercorrelation across five time points in Grades 6–8 ranged from 0.78–0.95. All analyses were conducted with the total number of correct targets identified in 3 min.

Test of Silent Reading Efficiency and Comprehension: The Test of Silent Reading Efficiency and Comprehension (TOSREC) is a nationally normed, group-administered test of silent reading of connected text for comprehension (Wagner, Torgesen, Rashotte, & Pearson, 2010). During standard administration, the student is required to read a series of short sentences and indicate whether the sentence is true or false. The raw score is the number of sentences correctly identified as true or false minus the incorrect responses within 3 min. If the number of incorrect responses exceeds the number of correct responses, a raw score of zero is recorded. Within our sample, the mean intercorrelation across five time points in Grades 6–8 ranged from 0.79–0.86 (Vaughn, Wanzek et al., 2010). We used age-based standard scores.

Analyses

Research Questions 1 and 2 were assessed using a split-plot design to compare group performance across the six cognitive variables. We followed procedures outlined by Huberty and Olejnik (2006) for a descriptive discriminant analysis to permit the interpretation of the contribution of specific dependent variables to the discriminant function (i.e., group separation). This design allows a simultaneous analysis of all variables. It answers questions pertaining to the effects of the grouping variable (responder status) on the set of outcome variables or, more specifically, to group separation on the outcome variables. This design is appropriate to answer Research Questions 1 and 2 because it addresses two issues: (a) whether groups differ across the set of outcome variables and (b) whether groups exhibit a distinct pattern (i.e., profile) of performance across the set of variables.

The analysis plan encompassed several steps. On the initial omnibus analysis, we first evaluated the Group-by-Task interaction to determine whether the effect of grouping variable (responder status) was consistent across the set of dependent variables. In the absence of an interaction, we evaluated main effects for group to determine whether groups differed on the set of dependent variables. Significant interactions and main effects were followed by pairwise multivariate comparisons of all possible group combinations to identify differences between adequate and inadequate responders and differences among inadequate responders identified through the application of different response criteria. This analysis permits interpretation of which specific groups (i.e., adequate responders and discrete inadequate responder groups) differ on the set of dependent variables. To control for a potential Type I error, a Bonferroni-adjusted α of $p < .008$ ($.05/6$) was used for all pairwise multivariate comparisons. Each pairwise comparison computes a linear discriminant function, which maximally separates the groups. Following procedures described by Huberty and Olejnik (2006), we report three methods for interpreting the contribution of specific variables to the discriminant function: canonical structure correlations, standardized discriminant function coefficients, and univariate contrasts. Univariate significance is evaluated at a Bonferroni-adjusted α of $p < .008$ ($.05/6$ to adjust for the six univariate contrasts). When only two groups are compared, univariate contrasts parallel the findings of canonical structure correlations but can be useful because there are no statistical tests associated with the two multivariate methods for interpreting the discriminant function (Huberty & Olejnik, 2006).

Research Question 3 was assessed following procedures outlined by Stanovich and Siegel (1994), who evaluated cognitive correlates of students with and without IQ-achievement discrepancies. These same procedures were used in a previous article investigating the cognitive and academic attributes of adequate and inadequate responders to an early elementary school reading intervention (Fletcher et al., 2011). Following these examples, we created six regression models, one model predicting each of the cognitive variables included in this report. The four predictor variables comprise the three response criterion measures (WJ-III Basic Reading, TOWRE, and WJ-III Passage Comprehension) and a contrast reflecting adequate or inadequate responder status. The contrast determines whether there is unique variance associated with the relation between performance on the cognitive variable and responder status beyond the variance explained by performance on the criterion reading

measures. Statistically significant β weights for the group contrast would suggest that the continuum-of-severity hypothesis (Vellutino et al., 2006) is insufficient to explain intervention responsiveness among adolescent readers.

RESULTS

We first investigated whether groups could be combined to maximize group size and reduce the number of comparisons. The comprehension and DFC groups were sufficiently large and theoretically important and were therefore left intact. However, the groups with specific deficits in fluency or decoding, as well as the groups falling below cut points in two of three criterion measures (i.e., the decoding and comprehension, decoding and fluency [DF], and fluency and comprehension [FC] groups), were too small to permit independent analyses, and differences in group assignment may reflect the measurement error of the tests. We therefore investigated whether the fluency, FC, and DF groups could be combined to form a group marked by fluency impairments. A MANOVA assessed whether the three groups performed differently on three measures of reading not used for group formation. Dependent variables included the GRADE reading comprehension standard score, AIMSweb Maze, and TOSREC standard score, and the independent variable was group membership (fluency, FC, and DF). The MANOVA was not statistically significant, $F(6, 80) = 1.06, p > .05, \eta^2 = 0.14$, suggesting the groups performed similarly in reading. We therefore combined the three groups into a single group marked by fluency impairments (hereafter called “the fluency group”; $n = 45$). The decoding and comprehension group and decoding group ($n = 8$ and $n = 8$, respectively) were too small to permit further analyses and were excluded from subsequent analyses. A MANOVA comparing excluded participants with remaining participants on the three external measures of reading was not significant, $F(3, 233) = 1.03, p > .05, \eta^2 = 0.01$.

Sociodemographic Variables

Table 1 provides mean age and frequency data for free and reduced-price lunch, history of English as a second language (ESL) status (all participating students were considered proficient and received instruction in English), and ethnicity for the four groups. There were significant differences in age across the four groups, $F(3, 217) = 16.01, p < .0001, \eta^2 = 0.18$. The DFC group was older than the comprehension, fluency, and responder groups, with mean age differences ranging from 0.53–0.86 years. For comparisons of cognitive data, this difference was addressed by using age-based standard scores when possible. We also evaluated relations between group status and other sociodemographic variables. There was a significant association between history of ESL status and group membership, $X^2(3, n = 215) = 8.06, p < .05$. Students in the poor comprehension group were more likely to have a history of ESL than students in the responder, DFC, and poor fluency groups. There was a significant association between special education status (identified for special education versus not identified for special education), $X^2(3, n = 215) = 40.86, p < .05$. Students in the DFC group were most likely to have been identified for special education, whereas the responder group was least likely to have been identified for special education. There was also a significant association between ethnicity and group membership, $X^2(9, n = 221) = 27.69, p < .05$, with a greater percentage of Hispanic students (18.5%) in the poor

comprehension group. The DFC group comprised a larger percentage of African American students. The association of group membership with gender, $X^2(3, n = 221) = 2.85, p > .05$, and free or reduced-lunch status, $X^2(3, n = 215) = 7.16, p > .05$, was not statistically significant.

Cognitive Variables

Means and standard deviations of the six cognitive measures for each group are presented in Table 3. A comparison of the z -score profiles for each group is presented in Figure 1. A split-plot design comparing the performance of the four groups on all six measures showed a significant Group-by-Task interaction, $F(15, 553) = 3.04, p < .0001, \eta^2 = 0.20$, with a moderate effect size. To investigate this interaction, we performed six pairwise multivariate comparisons investigating main effects and interaction terms.

Poor Comprehension Versus Responders—The interaction term for the comparison of the responder and poor comprehension groups was significant, $F(5, 119) = 5.44, p < .008, \eta^2 = 0.19$, with a moderate effect size. To help interpret the significant interaction, the discriminant structure coefficients, canonical structure correlations, and univariate contrasts are reported in Table 4. The three methods for interpreting the contribution of specific variables to the discriminant function maximally separating groups concurred in heavily weighting verbal knowledge and listening comprehension. The univariate contrast to phonological awareness was significant, and the standardized coefficients indicated a moderate contribution to the discriminant function. Processing speed, rapid naming, and nonverbal reasoning had comparatively small contributions to the discriminant function, and the univariate contrasts did not meet the critical level of α .

Poor Comprehension Versus Poor Fluency—The Group-by-Task interaction was significant for the poor fluency and poor comprehension group comparison, $F(5, 91) = 4.65, p < .001, \eta^2 = 0.20$, with a moderate effect size. Table 4 reports canonical correlations, standardized discriminant function coefficients, and univariate contrasts. The three methods indicated that verbal knowledge and listening comprehension contributed most to the discriminant function maximally separating groups. On both of these tasks, the poor fluency group scored significantly higher than the poor comprehension group. Rapid naming was also moderately weighted in its contribution to group separation, and the univariate contrast was significant, $p < .008$. However, on this task, the poor comprehension group performed better than the poor fluency group. Phonological awareness, processing speed, and nonverbal reasoning had comparatively small contributions to group separation, and the univariate contrasts were all nonsignificant.

Poor Comprehension Versus DFC—The Group-by-Task interaction was also significant for the poor comprehension versus DFC groups, $F(5, 86) = 4.49, p < .008, \eta^2 = 0.21$. Table 4 shows that phonological awareness contributed most to the discriminant function maximally separating groups. Rapid naming, processing speed, and nonverbal reasoning were also moderately weighted in their contribution to the discriminant function. In addition, the univariate contrasts for the three variables were significant. Listening

comprehension and verbal knowledge did not contribute meaningfully to the discriminant function, and both univariate contrasts were nonsignificant.

DFC Versus Responders—The pairwise multivariate comparisons of the responder and DFC groups showed no statistically significant interaction, $F(5, 105) = 1.86, p > .008, \eta^2 = 0.08$, with a small effect size. The main effect for group was significant, $F(6, 104) = 17.26, p < .001, \eta^2 = 0.50$, with a large effect. All of the univariate contrasts achieved the critical level of $\alpha, p < .008$. Table 4 shows that phonological awareness contributed most to the discriminant function. Rapid naming and verbal knowledge also contributed moderately. Listening comprehension, processing speed, and nonverbal reasoning contributed minimally to the discriminant function.

DFC Versus Poor Fluency—The pairwise multivariate comparisons of the poor fluency and DFC groups showed no significant interaction, $F(5, 75) = 0.72, p > .008, \eta^2 = 0.05$. The main effect for group was significant, $F(6, 76) = 6.04, p < .001, \eta^2 = 0.32$, with a large effect. Standardized discriminant function coefficients weighted verbal knowledge most heavily for group separation. Univariate contrasts for phonological awareness and nonverbal reasoning achieved the critical level of α and both variables correlate strongly with the canonical structure, but standardized discriminant function coefficients were weighted less heavily. Rapid naming and processing speed were not significant on univariate contrasts.

Poor Fluency Versus Responders—Pairwise multivariate comparisons of the responder and poor fluency groups showed no significant Group-by-Task interaction, $F(5, 110) = 1.96, p > .008, \eta^2 = 0.08$, or main effect, $F(6, 109) = 2.14, p > .008, \eta^2 = 0.13$, both with small to medium effects. No univariate contrasts achieved the critical level of α .

Regression Analyses: A Continuum of Severity?

To answer Research Question 3, we created regression models predicting the cognitive variables analyzed in this article. Each regression model consisted of four predictor variables: the three criterion measures used to determine responder status (WJ-III Basic Reading, TOWRE, and WJ-III Passage Comprehension) and a contrast reflecting adequate and inadequate responder status (dummy coded as 1 for inadequate responder and 0 for adequate responder). An analysis of the significance of the contrast determines whether there is unique variance in the cognitive variable associated with responder status beyond the variance explained by performance on the criterion reading measures. Such a finding would suggest that a continuum-of-severity hypothesis is inadequate for predicting intervention responder status and would provide support for the unique importance of cognitive assessment in adolescent struggling readers.

Across the seven models, the contrast for responder status was significant only for the model predicting nonverbal reasoning, $b = 0.27, t(220) = 1.70, p < .05$. The positive sign of the b weight adjusts the predicted mean of the nonverbal reasoning score of inadequate responders higher than would be predicted given their performance on the three criterion measures. The addition of the contrast resulted in an increase in explained variance from 9.8% to 11.3%.

The contrast of responder versus inadequate responder did not explain significant unique variance in any of the other models, consistent with a continuum-of-severity hypothesis.

DISCUSSION

The first research question addressed whether there are cognitive attributes that differentiate inadequate and adequate responders to a Tier 2 intervention. Our results suggest that adequate and inadequate responders can be differentiated across cognitive variables because contrasts with the adequate responder groups were largely significant. Group separation is apparent in Figure 1, where the adequate responder group presents a flatter, generally higher profile than the inadequate responder groups, who show uneven performance with specific deficits related to documented reading deficits. This finding provides evidence for the validity of inadequate and adequate responder status as a classification attribute because resulting groups can be differentiated on variables not used for group formation (Morris & Fletcher, 1998).

The second question addressed whether inadequate responder groups could be differentiated across cognitive attributes based on the assessed reading domains. The results of our study suggest that in middle school, it is possible to identify at least three groups of inadequate responders in addition to an adequate responder group. Each group showed unique cognitive skill profiles, consistent with previous research investigating the cognitive profiles of good and poor readers defined according to decoding, fluency, and comprehension criteria.

Cognitive Correlates of Intervention Responder Status

The Group-by-Task interactions on cognitive measures (visually presented in Figure 1) were striking. On each pairwise multivariate comparison of cognitive skills that included the comprehension group, there was a significant Group-by-Task interaction with effect sizes in the moderate to large range. This effect is clearly illustrated in Figure 1, in which the performance of the comprehension group drops sharply on the listening comprehension and verbal knowledge tasks. On both of these tasks, the performance of the comprehension group is not significantly different from that of the generally lower performing DFC group but is significantly lower than that of the responder and fluency groups.

The strong role of listening comprehension and verbal knowledge in group separation in comparisons including the comprehension group is not unexpected. Although previous multivariate analyses of the cognitive correlates of inadequate response have not found a strong contribution of oral language toward group separation (Fletcher et al., 2011; Stage et al., 2003; Vellutino et al., 2006), our study included a reading comprehension criterion measure, which may have identified previously unidentified inadequate responders. Second, our sample included older students. As students age, the cognitive demands of the reading task change. In early elementary school, limits in word reading accuracy and fluency limit the complexity of text, and most young readers are able to comprehend successfully decoded text with little difficulty. However, as students progress through school, academic text becomes more complex and cognitively taxing. Many students with limited academic language and vocabulary may begin to show comprehension difficulties in late elementary school (Catts et al., 2005; Chall, Jacobs, & Baldwin, 1990; Lesaux & Kieffer, 2010).

Previous investigations of students with comprehension impairments in middle school also found a close connection between listening comprehension and vocabulary and impairments in reading comprehension (Catts et al, 2006; Lesaux & Kieffer, 2010). These findings lend support to the existence of a distinct subtype of reading disability, marked by specific impairments in reading comprehension, with language and vocabulary deficits implicated as correlates. This subgroup is more apparent in older students.

It is also noteworthy that there was a statistically significant association between ESL status and comprehension group membership. ESL students were more likely to be identified as inadequate responders according to comprehension criteria. Previous investigations of specific comprehension difficulties among monolingual students in late elementary school suggested that only a small percentage of students show specific comprehension deficits (Catts et al, 2005). However, the prevalence of specific comprehension deficits may be more common among English language learners (ELLs) because of relative skill deficits in vocabulary and listening comprehension (Jean & Geva, 2009; Mancilla-Martinez & Lesaux, 2010). Similar to monolingual students, these deficits in oral language skill may manifest in late emerging, specific comprehension difficulties (Nakamoto, Lindsey, & Manis, 2007). Such findings highlight the uniquely challenging task facing ELLs trying to acquire grade-level reading proficiency in their second language. Such findings highlight a need for ongoing vocabulary and oral language instruction for ELLs into middle school, a longer duration than may be common.

In contrast to comparisons involving the comprehension group, comparisons including the poor fluency and DFC groups implicated phonological awareness as a significant contributor to group separation. This is consistent with previous investigations of inadequate responders conducted in younger readers (Fletcher et al., 2011; Stage et al., 2003; Vellutino et al., 2003). However, our findings differ from the findings of Stage et al. (2003) in the more limited role of rapid naming in group separation. In comparisons of the poor fluency and DFC groups with the adequate responder group, rapid naming was weighted less heavily than phonological awareness. This is consistent with previous research suggesting that the relation of rapid automatized naming to reading outcomes shows differences over time (Wagner et al., 1997). Although the present multivariate analyses did not find a large, unique contribution of rapid naming to group separation, it is important to acknowledge the sharp drop in rapid naming for the fluency group, consistent with previous research investigating the characteristics of fluency-impaired adolescents (Barth et al., 2009).

Continuum of Severity

The third research question addressed whether there was unique variance in cognitive performance associated with responder status beyond that explained by performance on the three criterion measures of reading. Only the contrast in the model predicting nonverbal reasoning accounted for unique variance beyond that explained by the three criterion measures. The amount of explained variance was small (1.4%) and not in a direction that would support the added value of cognitive assessment following inadequate RTI. The model adjusted the performance of inadequate responders on the nonverbal reasoning task higher than would have been predicted by performance on the criterion measures, narrowing

the gap in performance between inadequate and adequate responders. If inadequate RTI is associated with unique cognitive deficits beyond those accounted for in reading performance, we would expect that more of these contrasts would be significant. Such a finding would suggest a need to assess cognitive processes following a determination of inadequate response to identify specific cognitive deficits that may be impairing academic progress in ways unobservable through academic assessment alone. However, the findings of our study do not support this notion. Instead, differences in cognitive performance paralleled differences in academic performance, suggesting little value added through cognitive assessment. The results of the regression analyses are consistent with the continuum-of-severity hypothesis, which posits that differences in the cognitive profiles of good and poor readers can be explained by differences in the severity of reading impairment (Fletcher et al., 2011; Vellutino et al., 2003, 2006).

Limitations of Study

The results of the study are limited to the sample, intervention approach, and methods described. This is particularly salient because there is a dearth of research investigating RTI in middle school. For example, the intervention provided was longer (45–50 min) than is typical in elementary school studies of RTI. In addition, the amount of training and support provided to intervention teachers may not align with typical practice in schools. Regarding our sample, we should draw specific attention to our criteria for adequate response. We used cut points on three psychometrically strong standardized assessments of reading. Different measures would change group membership and may have affected the results. However, this fact also highlights the necessity of using multiple assessment measures. No single measure would have provided adequate coverage to identify all students who continue to need intensive reading interventions.

Another important consideration is the number of adequate responders to the Tier 2 intervention. Roughly one in three participants met multiple criteria for adequate response. This likely reflects the difficulty of remediating reading difficulties at older ages and the use of multiple response criteria. However, it is possible that the intervention was insufficiently intensive or not ideally matched to the educational needs of poor readers in middle school.

Limitations also arise from the design of the larger study. For example, the KBIT-2 Matrix Reasoning and KBIT-2 Verbal Knowledge tests were administered at different times than other cognitive measures, which was reasonable because standard scores were used and these tests are stable over a school year. In addition, the criterion reading measures were administered in spring of Year 1, whereas the cognitive tests (except the KBIT-2 subtests) were administered in fall of Year 2. This reflected the constraints of a large-scale intervention study and the need to limit the amount of assessment at any one time. It is difficult to determine the effects of these disparate testing times and any potential effects caused by a “summer slump,” which has not been studied extensively among adolescents. However, we note that disparate testing times only affect cognitive measures, which are theoretically more stable than academic measures. Testing time was limited, necessitating these decisions.

Although this study found a significant Group-by-Task interaction when comparing cognitive attributes of groups of adequate and inadequate responders, it was not designed to investigate potential Aptitude-by-Treatment interactions. We have found evidence for distinct cognitive correlates for different groups of inadequate responders, but it should not be inferred that such differences necessitate different approaches to reading intervention based on cognitive functioning because the variability was accounted for by variations in the pattern of reading difficulties.

Implications for Practice

The results of this study highlight the importance of using multiple measures across reading domains to determine adequate RTI. The use of any single criterion measure in this study would have resulted in a much larger number of students identified as adequate responders. In schools this may result in a large number of students being ineligible for necessary intervention, despite the need documented by a more comprehensive evaluation of their reading skills. Through the assessment of multiple domains of reading, we were able to identify discrete groups with specific reading deficits in fluency and comprehension. This comprehensive evaluation of reading skill is even more important at the middle school level, for which there is a dearth of psychometrically validated curriculum-based measures that could be used to evaluate growth or dual-discrepancy models.

The between-group differences in performance on the criterion reading measures also have implications for intervention design. To the extent that intervention should be tailored to the needs of individual students or groups of students, the results of this study would suggest that the simplest and most efficient method would be to vary instruction to target specific academic deficits, rather than matching instructional style or content to specific cognitive deficits. In the context of reading intervention, Connor et al. (2009) documented promising outcomes for treatments tailored to the reading needs of individual students (i.e., meaning-based instruction versus code-based instruction). In contrast, Aptitude-by-Treatment interactions based on cognitive processes remain largely unproven and speculative (Kearns & Fuchs, 2013; Pashler, McDaniel, Rohrer, & Bjork, 2009). In the absence of compelling evidence for Aptitude-by-Treatment interactions, practitioners would be better served by matching instruction to academic need.

The clear separation between the adequate and inadequate responder groups on the set of cognitive measures has implications for the identification of LD. There is no gold standard by which LD classification decisions can be evaluated. In the absence of a gold standard, potential classification systems must be validated through comparisons of resulting groups. If resulting groups show meaningful differences on external variables, such as cognitive, behavioral, or neurobiological attributes, the classification system would accrue validity (Morris & Fletcher, 1998). Our finding that adequate and inadequate responders can be consistently differentiated across a set of cognitive variables not used for group formation provides important evidence for the validity of intervention responder status as a potential classification attribute of struggling readers in middle school.

CONCLUSIONS

We investigated the cognitive attributes of inadequate and adequate responders to intervention. We identified three groups of inadequate responders with specific difficulties in comprehension, fluency, and more global reading difficulties. The three inadequate responder groups and the adequate responder group showed distinct cognitive profiles, with relations between cognitive and academic deficits following theoretically predicted patterns and replicating previous research with middle school readers. However, the results of this study suggest that these differences in cognitive functioning can be largely explained based on the correlations of different academic and cognitive variables. If, as our study suggests, the results of a comprehensive cognitive assessment following inadequate RTI do not explain responder status or inform subsequent interventions, this raises questions about the utility of comprehensive cognitive assessment following a determination of inadequate response. Our findings suggest that limited resources may be better used in directly assessing and teaching important academic skills.

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; 23:300–316.
- Archer, AL., Gleason, MM., Vachon, V. REWARDS intermediate: Multisyllabic word reading strategies. Sopris West; Longmont, CO: 2005a.
- Archer, AL., Gleason, MM., Vachon, V. REWARDS plus: Reading strategies applied to social studies passages. Sopris West; Longmont, CO: 2005b.
- Barth AE, Catts HW, Anthony JL. The component skills underlying reading fluency in adolescent readers: A latent variable analysis. *Reading and Writing*. 2009; 22:567–590.
- Barth AE, Stuebing KK, Anthony JL, Denton CA, Mathes PG, Fletcher JM, Francis DJ. Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences*. 2008; 18:296–307. [PubMed: 19081758]
- 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; 32:557–582.
- Catts HW, Adlof SM, Weismar SE. Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*. 2006; 49:278–293.
- Catts, HW., Hogan, TP., Adlof, SM. Developmental changes in reading and reading disabilities. In: Catts, HW., editor. *The connections between language and reading disabilities*. Lawrence Erlbaum Associates; Mahwah, NJ: 2005. p. 25-40.
- Chall, JS., Jacobs, VA., Baldwin, LE. *The reading crisis: Why poor children fall behind*. Harvard University Press; Cambridge, MA: 1990.
- Cirino PT, Romain MA, Barth AE, Tolar TD, Fletcher JM, Vaughn S. Reading skill components and impairments in middle school struggling readers. *Reading and Writing*. 2012; 26:1059–1086.
- Compton D, Fuchs LS, Fuchs D, Lambert W, Hamlett C. The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities*. 2012; 45:79–95. [PubMed: 21444929]
- Connor CM, Piasta S, Fishman B, Glassney S, Schatschneider C, Crowe E, Morrison FJ. Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Development*. 2009; 80:77–100. [PubMed: 19236394]
- Doehring, DG. *Patterns of impairment in specific reading disability*. University Press; Bloomington, IN: 1968.

- Evans JJ, Floyd RG, McGrew KS, Leforgee MH. The relations between measures of Cat-tell-Hom-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review*. 2001; 31:246–262.
- Fiorello CA, Hale JB, Snyder LE. Cognitive hypothesis testing and response to intervention for children with reading problems. *Psychology in the Schools*. 2010; 43:835–853.
- Fletcher JM, Stuebing KK, Barth AE, Denton CA, Cirino PT, Vaughn S. Cognitive correlates of inadequate response to reading intervention. *School Psychology Review*. 2011; 40:3–22. [PubMed: 23125475]
- Fletcher JM, Vaughn S. Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*. 2009; 3:30–37. [PubMed: 21765862]
- Fuchs D, Deshler DK. What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disability Quarterly*. 2007; 27:216–227.
- Gresham FM, Reschly DJ, Tilly WD, Fletcher J, Bums M, Crist T, Shinn MR. A response to intervention perspective. *The School Psychologist*. 2005; 59:26–33.
- Hale J, Alfonso V, Beminger B, Bracken B, Christo C, Clark E, Yalof J. Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly*. 2010; 33(3):223–236.
- Huberty, CJ., Olejnik, S. *Applied discriminant analysis*. 2nd. Wiley; New York, NY: 2006.
- Jean M, Geva E. The development of vocabulary in English as a second language children and its role in predicting word recognition ability. *Applied Psycholinguistics*. 2009; 30:153–185.
- Lesaux NK, Kieffer MJ. Exploring sources of reading comprehension difficulty among language minority learners and their classmates in early adolescence. *American Educational Research Journal*. 2010; 47:596–632.
- Lipsey, MW., Puzio, K., Yun, C., Hebert, MA., Steinka-Fry, K., Cole, MW., Roberts, M., Busick, MD. Translating the statistical representation of the effects of education interventions into more readily interpretable forms (NCSE 2013-3000). National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education; Washington, DC: 2012. Retrieved from Institute of Education Sciences website: <http://ies.ed.gov/ncser/>
- Kaufman, AS., Kaufman, NL. *Kaufman brief intelligence test*. 2nd. Pearson Assessment; Minneapolis, MN: 2004.
- Kearns DM, Fuchs D. Does cognitively focused instruction improve the academic performance of low-achieving students? *Exceptional Children*. 2013; 79:263–290.
- Mancilla-Martinez J, Lesaux NK. Predictors of reading comprehension for struggling readers: The case for Spanish-speaking language minority learners. *Journal of Educational Psychology*. 2010; 102:701–711. [PubMed: 20856691]
- McGrew, KS. *Journal of Psychoeducational Assessment monograph series: Woodcock-Johnson psycho-educational assessment battery—Revised*. Psychoeducational; Cordova, TN: 1983. The relationship between the Woodcock-Johnson Psycho-Educational Assessment Battery—Revised *Gf-Gc* cognitive clusters and reading achievement across the life-span; p. 39-53.
- McMaster KL, Fuchs D, Fuchs LS, Compton DL. Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*. 2005; 71:445–463.
- Morris RD, Fletcher JM. Classification in neuropsychology: A theoretical framework and research paradigm. *Journal of Clinical and Experimental Neuropsychology*. 1998; 10:640–658.
- Nakamoto J, Lindsey KA, Manis FR. A longitudinal analysis of English reading comprehension. *Reading and Writing*. 2007; 20:691–719.
- Nelson RJ, Benner GJ, Gonzalez J. Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice*. 2003; 18:255–267.
- Pashler H, McDaniel M, Rohrer D, Bjork R. Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*. 2009; 9(3):105–119.
- Scammacca N, Roberts G, Vaughn S, Stuebing KK. A meta-analysis of interventions for struggling readers in grades 4-12, 1980-2011. *Journal of Learning Disabilities*. in press.

- Schatschneider C, Wagner RK, Crawford EC. The importance of measuring growth in response to intervention models: Testing a core assumption. *Learning and Individual Differences*. 2008; 18:308–315. [PubMed: 22224065]
- Shinn, MR., Shinn, MM. AIMSweb training workbook: Administration and scoring of reading maze for use in general outcome measurement. Edformation; Eden Prairie, MN: 2002.
- Stage SA, Abbott RD, Jenkins JR, Berninger VW. Predicting response to early reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ-word reading discrepancy: Failure to validate discrepancy method. *Journal of Learning Disabilities*. 2003; 36:24–33. [PubMed: 15490889]
- Stanovich KE. Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly*. 1991; 26:7–29.
- Stanovich KE, Siegel LS. Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology*. 1994; 86:24–53.
- Tolar TD, Barth AE, Fletcher JM, Francis DJ, Vaughn S. Predicting reading outcomes with progress monitoring slopes among middle grade students. *Learning and Individual Differences*. 2014; 30:46–57. [PubMed: 24659899]
- Torgesen, J., Wagner, R., Rashotte, C. Test of word reading efficiency. Pro-Ed; Austin, TX: 1999.
- VanDerHeyden, AM., Bums, MK. Essentials of response to intervention. John Wiley & Sons; Hoboken, NJ: 2010.
- Vaughn S, Cirino PT, Wanzek J, Wexler J, Fletcher JM, Denton CD, Francis DJ. Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review*. 2010; 39:3–21. [PubMed: 21479079]
- Vaughn S, Wanzek J, Wexler J, Barth A, Cirino PT, Fletcher JM, Francis DJ. The relative effects of group size on reading progress of older students with reading difficulties. *Reading and Writing*. 2010; 23:931–956. [PubMed: 21072131]
- Vaughn S, Wexler J, Leroux A, Roberts G, Denton C, Barth A, Fletcher JM. Effects of intensive reading intervention for eighth-grade students with persistently inadequate response to intervention. *Journal of Learning Disabilities*. 2012; 45:515–525. [PubMed: 21512102]
- Vaughn S, Wexler J, Roberts G, Barth A, Cirino P, Romain M, Denton CA. Effects of individualized and standardized interventions on middle school students with reading disabilities. *Exceptional Children*. 2011; 77:391–407. [PubMed: 23125463]
- Vellutino, FR., Scanlon, DM., Jaccard, J. Toward distinguishing between cognitive and experiential deficits as primary sources of difficulty in learning to read: A two-year follow-up to difficult to remediate and readily remediated poor readers. In: Foorman, BR., editor. Preventing and remediating difficulties. York Press; Baltimore, MD: 2003. p. 73-120.
- Vellutino FR, Scanlon DM, Small S, Fanuele DP. Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities*. 2006; 39:157–169. [PubMed: 16583795]
- Wagner, RK., Torgesen, JK., Rashotte, CA. The comprehensive test of phonological processes (CTOPP). PRO-ED; Austin, TX: 1999.
- Wagner RK, Torgesen JK, Rashotte CA, Hecht SA, Barker TA, Burgess SR, Garon T. Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A five year longitudinal study. *Developmental Psychology*. 1997; 33:468–479. [PubMed: 9149925]
- Wagner, RK., Torgesen, JK., Rashotte, CA., Pearson, NA. Test of sentence reading efficiency and comprehension (TOSREC). PRO-ED; Austin, TX: 2010.
- Williams, KT. The group reading assessment and diagnostic evaluation (GRADE): Teacher's scoring and interpretive manual. American Guidance Service; Circle Pines, MN: 2001.
- Wolff P. Impaired temporal resolution in developmental dyslexia: Temporal information processing in the nervous system. *Annals of the New York Academy of Sciences*. 1993; 682:87–103. [PubMed: 7686728]

Woodcock, RW., McGrew, KS., Mather, N. Woodcock-Johnson HI tests of achievement. Riverside; Itasca, IL: 2001.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Cognitive Profiles of Adequate and Inadequate Responders

Note. C = comprehension-impaired group; DFC = decoding, fluency, and comprehension-impaired group; F = fluency-impaired group; Responder = adequate responder group; Phonological Awareness = Comprehensive Test of Phonological Processing (CTOPP) Composite; Listening Comprehension = Group Reading Assessment and Diagnostic Evaluation (GRADE) Listening Comprehension; Rapid Naming = CTOPP Rapid Automated Naming-Letters; Processing Speed = Underlining Test; Nonverbal Reasoning = Kaufman Brief Intelligence Test–Second Edition (KBIT-2) Matrix Reasoning subtest; Verbal Knowledge = KBIT-2 Verbal Knowledge subtest.

Table 1

Demographic Data by Responder Group

Variable	C (<i>n</i> = 54)	F (<i>n</i> = 45)	DFC (<i>n</i> = 45)	Responder (<i>n</i> = 77)
Age, years *				
<i>M</i>	11.75	11.7	12.28	11.42
<i>SD</i>	0.77	0.64	0.8	0.49
% Male	59.26	46.67	53.33	45.45
% Free or reduced-price lunch	90.74	75	87.5	75.32
% ESL *	18.52	11.36	7.5	3.9
% Special education *	12.96	20.45	56.10	7.79
Race or ethnicity, % *				
Black	27.78	44.44	64.44	45.45
White	3.7	13.33	4.44	18.18
Hispanic	64.81	42.22	31.11	35.06
Other	3.7	0	0	1.3

Note. C = comprehension impaired; DFC = decoding, fluency, and comprehension impaired; ESL = English as a second language; F = fluency impaired.

* $p < .05$

Table 2

Means and Standard Deviations for Inadequate Responders by Criteria Group in Original Measurement Units

Criterion Measure	C (n = 54)		D (n = 8)		DC (n = 8)		DF (n = 11)		DFC (n = 45)		F (n = 19)		FC (n = 15)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
WJ-III Basic Reading	100.06	7.94	88.50	1.69	87.25	2.43	87.00	2.97	80.73	6.70	96.58	4.69	97.53	6.56
TOWRE	104.56	10.19	100.75	7.09	96.00	3.63	84.55	3.86	77.24	8.70	85.26	4.69	85.07	4.70
WJ-III Passage Comprehension	84.35	5.19	96.38	3.11	83.00	6.02	94.55	3.72	77.38	9.39	96.84	5.46	84.53	4.91

Note. $M = 100$, $SD = 15$ on all measures. C = impaired on Passage Comprehension only; D = impaired on Basic Reading only; DC = impaired on Basic Reading and Passage Comprehension; DF = impaired on Basic Reading and TOWRE; DFC = impaired on Basic Reading, TOWRE, and Passage Comprehension; F = impaired on TOWRE only; FC = impaired on TOWRE and Passage Comprehension; TOWRE = Test of Word Reading Efficiency; WJ-III = Woodcock-Johnson III.

Table 3

Performance by Group on Cognitive Measures in Original Measurement Units

Cognitive Variable	Comprehension (<i>n</i> = 54)		Fluency (<i>n</i> = 45)		DFC (<i>n</i> = 45)		Responder (<i>n</i> = 11)	
	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>
Phonological awareness ^a	8.72	2.15	8.52	2.76	6.82	2.16	9.73	2.30
Listening comprehension ^b	9.04	2.01	10.47	2.24	9.39	2.21	10.75	2.15
Rapid naming ^c	8.85	3.71	7.22	3.07	6.18	2.60	8.43	2.83
Processing speed ^d	0.11	0.66	0.07	0.68	−0.35	0.65	0.09	0.65
Nonverbal reasoning ^e	96.50	11.03	98.67	10.32	90.67	13.37	97.55	10.22
Verbal knowledge ^f	81.92	9.31	90.02	11.35	80.78	15.85	91.23	10.67

Note. DFC = decoding, fluency, and comprehension impaired.

^aComprehensive Test of Phonological Processing-Phonological Awareness Composite (standard score; *M* = 10, *SD* = 3).

^bGroup Reading Assessment and Diagnostic Evaluation Listening Comprehension (stanine score).

^cComprehensive Test of Phonological Processing Rapid Automatized Naming–Letters subtest (standard score; *M* = 10, *SD* = 3).

^dAverage *z* score on three underlining tests.

^eKaufman Brief Intelligence Test Matrix Reasoning (standard score).

^fKaufman Brief Intelligence Test Verbal Knowledge (standard score).

Canonical Structure Correlations and Standardized Coefficients for Cognitive Dimensions of Group Status

Table 4

Cognitive Measure	Responder Versus C		Responder Versus F		Responder Versus DFC		C Versus F		C Versus DFC		F Versus DFC	
	r_s	<i>SDFC</i>	r_s	<i>SDFC</i>	r_s	<i>SDFC</i>	r_s	<i>SDFC</i>	r_s	<i>SDFC</i>	r_s	<i>SDFC</i>
Phonological awareness	0.46*	0.25	0.72*	0.82	0.81*	0.92	-0.12	-0.33	0.73*	0.73	0.61*	0.44
Listening comprehension	0.75*	0.56	0.14	0.1	0.38*	0.13	0.7*	0.57	-0.17	-0.36	0.41	0.1
Rapid naming	-0.17	-0.14	0.51	0.71	0.5*	0.48	-0.49	-0.42	0.62*	0.59	0.31	0.26
Processing speed	-0.05	-0.06	-0.01	-0.15	0.38*	0.2	-0.04	0.2	0.5*	0.46	0.5	0.38
Nonverbal reasoning	0.1	-0.08	-0.12	-0.35	0.46*	0.17	0.19	0.14	0.46*	0.12	0.62*	0.39
Verbal knowledge	0.84*	0.72	0.17	0.33	0.55*	0.41	0.77*	0.61	0.11	0.11	0.6*	0.64

Note. The comparison of the responder and fluency-impaired groups did not achieve the critical α ($p < .008$). C = comprehension impaired; DFC = decoding, fluency, and comprehension impaired; F = fluency impaired; *SDFC* = standardized discriminant function coefficient.

* Univariate $p < .008$ (.05/6).