

Research Bank

Journal article

Is "response/no response" too simple a notion for RTI frameworks? Exploring multiple response types with latent profile analysis

Peng, Peng, Fuchs, Douglas, Fuchs, Lynn S., Cho, Eunsoo, Elleman, Amy M., Kearns, Devin M., Patton III, Samuel and Compton, Donald L.

This is an Accepted Manuscript of an article published as:

Peng, P., Fuchs, D., Fuchs, L. S., Cho, E., Elleman, A. M., Kearns, D. M., Patton III, S. and Compton, D. L. (2020). Is "response/no response" too simple a notion for RTI frameworks? Exploring multiple response types with latent profile analysis. *Journal of Learning Disabilities*, 53(6), pp. 454-468. <https://doi.org/10.1177/0022219420931818>

This work © 2020 is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)



Published in final edited form as:

J Learn Disabil. 2020 ; 53(6): 454–468. doi:10.1177/0022219420931818.

Is “Response/No Response” Too Simple a Notion for RTI Frameworks? Exploring Multiple Response Types with Latent Profile Analysis

Peng Peng, Ph.D¹, Douglas Fuchs, Ph.D², Lynn S. Fuchs, Ph.D², Eunsoo Cho, Ph.D³, Amy M. Elleman, Ph.D⁴, Devin M. Kearns, Ph.D⁵, Samuel Patton III, Ph.D², Donald L. Compton, Ph.D⁶

¹University of Texas-Austin

²Vanderbilt University

³Michigan State University

⁴Middle Tennessee State University

⁵University of Connecticut

⁶Florida State University

Abstract

We conducted a secondary analysis of data from a randomized control trial to explore this question: Does “response/no response” best characterize students’ reactions to a generally efficacious 1st-grade reading program, or is a more nuanced characterization necessary? Data were collected on 265 at-risk readers’ word reading prior to and immediately following program implementation in 1st grade and in spring of 2nd grade. Pretreatment data were also obtained on domain-specific skills (letter knowledge, decoding, passage comprehension, language) and domain-general skills (working memory, non-verbal reasoning). Latent profile analysis of word reading across the 3 time points with controls as a local norm revealed a *strongly responsive* group ($n = 45$) with mean word reading z scores of .25, 1.64, and 1.26 at the 3 time points, respectively; a *mildly responsive* group ($n = 109$), z scores = .30, .47, and .55; a *mildly non-responsive* group ($n = 90$), z scores = -.11, -.15, and -.55; and a *strongly non-responsive* group ($n = 21$), z scores = -1.24, -1.26, and -1.57. The 2 responsive groups had stronger pretreatment letter knowledge and passage comprehension than the 2 non-responsive groups. The mildly non-responsive group demonstrated better pretreatment passage comprehension than the strongly non-responsive group. No domain-general skill distinguished the 4 groups. Findings suggest response to early reading intervention was more complicated than response/no response, and pretreatment reading comprehension was an important predictor of response even with pretreatment word reading controlled.

Corresponding authors: Peng Peng, kevpp2004@hotmail.com; Department of Special Education, College of Education, The University of Texas at Austin, SZB 408D, 1912 Speedway STOP D5300, Austin, TX 78712 or Douglas Fuchs, doug.fuchs@vanderbilt.edu. Vanderbilt University PMB 228, Peabody College, Nashville, TN 37203-5721.

Keywords

latent profile analysis; at-risk readers; reading comprehension; domain-specific and domain-general skills; RTI

It has been estimated that between 5-10% of primary-grade students are at risk for serious reading problems (e.g., Wanzek & Vaughn, 2009), including many children with disabilities (e.g., Gilmour, Fuchs, & Wehby, 2019). Weak beginning reading skills affect children's enjoyment of reading, learning in content areas like mathematics and science, and social adjustment (e.g., Swanson, Harris, & Graham, 2013). Moreover, early reading difficulty may reverberate throughout one's educational career (Foorman, Francis, Shaywitz, Shaywitz, & Fletcher, 1997) and is associated with undesirable post-school outcomes like low income and poor health (Miller, McCardle, & Hernandez, 2010). That said, initial signs of reading problems need not sentence a student to chronic school failure and life-long misery. Proper early instruction can reduce the number of students with persistent reading problems (e.g., Torgesen, Alexander, Wagner, Rashotte, Voeller, & Conway, 2001).

Instructional Response

One of many challenges facing those who try to conduct effective early reading intervention is that at-risk readers do not respond similarly to a given academic program, including those shown to have benefitted many students. Response-to-intervention (RTI) research in the past 15 years or so has documented this fact time and again (cf. Tran, Sanchez, Arellano, & Swanson, 2011). At-risk students' dissimilar responses are all the more believable and important because RTI research typically engages these children in scientifically-validated programs, which is to say that a child's weak or inadequate response is unlikely attributable to weak or inadequate curricula and instruction. Rather, in RTI frameworks, inadequate response is viewed as a sign of a possible disorder, the probability of which becomes stronger with repeated unsatisfactory responses in more intensive instructional tiers (cf. Vaughn et al., 2010).

RTI frameworks—with their signature tiers of more and more intensive instruction—have become an increasingly popular approach to both disability identification and treatment (e.g., D. Fuchs & Fuchs, 2006). In some states, they have become so popular that instructional response has usurped IQ-achievement discrepancy as a principle means of learning disabilities (LD) identification (cf. North Carolina Department of Public Instruction, 2017). Despite such acceptance—or maybe because of it—we believe it is appropriate to explore how response to instruction is commonly understood by researchers and practitioners. To be clear about our purposes and research methods described below, we first situate response in a broader context.

Categorical and Dimensional Perspectives

Explaining Terms

Fletcher and Miciak (2019) suggest that we may generally think of disorders as categorical or dimensional in nature. Categorical disorders, they say, are often binary conditions. An

individual either has a disorder or doesn't. Most cancers and viruses, say Fletcher and Miciak, are categorical disorders. Other disorders are best understood dimensionally; as existing on a continuous distribution. An example of a dimensional disorder is high blood pressure. Fletcher and Miciak write that no natural threshold—no inherent or *a priori* cut-point—separates individuals with high versus low blood pressure. Instead, a line is drawn to identify hypertensive people based on empirically determined outcomes for this subgroup of the population.

Likewise, in a classroom or grade level, response to instruction may be ordered without division or interruption from weakest to strongest and understood dimensionally; that is, as differing in degree but not in kind. According to Fletcher, Lyons, Fuchs, and Barnes (2019), research suggests that classifications based on instructional response create subgroups that differ on a variety of theoretically related attributes including academic achievement and cognitive processes. This, they say, is evidence of the validity of instructional response as a classification attribute. They caution, however, that such subgroup differences are *not* evidence of a categorical disorder. Rather the differences lie on a continuum reflecting the severity of academic difficulty.

Competing or Complementary Approaches?

The subgroups in the research described by Fletcher et al. (2019) indicate something else as well: A dimensional view does not preclude dividing distributions into two or more categories to better understand the variability they reflect. That is, a dimensional view may be explored by means of creating categories. Similarly, a categorical approach may be supplemented by dimensional analyses. Stein (2012) elaborates on this complementary relationship when he writes, “Both the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* and the *International Classification of Disease (ICD)* systems rely...on a categorical approach, but [each] also note[s] the dimensional nature of syndromes and symptoms. [Whereas] in both systems, the main focus is on describing...categorical clinical entities [,] [the two systems also] use a range of specifiers, such as ‘mild,’ ‘moderate,’ and ‘severe,’ which introduce a dimensional aspect” (p. 270). As for dimensional disorders, Stein writes, “they are often translated back into categorical approaches via the use of cut points. [P]atients with scores above a particular cut point on a symptom severity rating scale may be viewed as having the relevant disorder, while those with scores below this cut point may be viewed as not having the condition” (p. 270).

Thinking Differently about Response

Apart from whether categorical and dimensional perspectives should or should not be considered mutually exclusive of each other, the use of instructional response has important limitations. For example, Fletcher and Miciak (2019) write that when response is understood (dimensionally) as a point on a continuum, there will be “inadequate” and “adequate” performers whose scores are close to the cut-point and, as a consequence, these designations will be unreliable (a limitation that applies to dimensional and categorical approaches alike). A related admonition is that a universally accepted definition of response does not exist (e.g., Barth et al., 2008; D. Fuchs, Fuchs, Compton, Bryant, & Davis, 2008; D. Fuchs,

Fuchs, & Compton, 2004). Rather, its definition and operationalization vary across states and districts as well in research investigating RTI (Fletcher & Miciak, 2019).

A less frequently discussed concern is that researchers and practitioners typically describe instructional response as a binary phenomenon with adequate response often defined as “program participants demonstrate greater pre-to-posttreatment reading growth than control students”; inadequate response as “program participants do no better than controls from pre-to-posttreatment.” Implicit in these definitions is another possible concern: Response/no-response often turns on performance that immediately follows completion of an intervention, reflecting little concern about whether program effects are sustained (cf. Wanzek, Stevens, Williams, Scammacca, Vaughn, & Sargent, 2019). In short, a binary view of response may oversimplify a complex set of responses to instruction. Put differently, response/no response ignores the possibility of *subgroups* of responders and non-responders, and it typically does not consider the maintenance of instructional effects (Wanzek et al., 2019).

In this paper, we explore the possibility of a more complex or nuanced response among at-risk students to an empirically-validated early reading program in hopes of providing information that enriches and refines an understanding of response that in turn leads to more effective instruction. Such exploration involves the use of categories to understand the variability often characterizing students’ response (which is consistent with school practice, whereby response/no-response determines the movement of students across the instructional tiers in RTI frameworks). But, as explained, this should not be interpreted to mean that we are pursuing a “categorical approach,” or that we are embracing this meta-view of disorder.

A Typology of Response to Early Intervention

Following is a description of a relatively complex response framework that we derived from our review of program evaluations in the early intervention literature; specifically, evaluations of the short- and long-term effects of instructional programs on young children’s cognitive and academic development, including their reading development (Bailey, Duncan, Odgers, & Yu, 2017; Bailey, Nguyen, Jenkins, Domina, Clements, & Sarama, 2016; Barnett, 2011; Ehrhardt, Huntington, Molino, & Barbaresi, 2013; Holmes, Gathercole, & Dunning, 2009; Protzko, 2015; Vandell, Belsky, Burchinal, Steinberg, Vandergrift, & NICHD Early Child Care Research Network, 2010). These early intervention outcomes may be described in terms of four response types that reflect short- and long-term effects. Explanations of their plausibility and usefulness follow.

Response-Response and Response-No Response

A first response category, *response-response*, refers to young children who show superior outcomes to controls both immediately following program completion and at follow up (e.g., Ehrhardt et al., 2013). One explanation of response-response is that many strong classroom-based early reading programs attempt to teach relatively accessible skills such as phonological awareness and beginning decoding, prerequisites for reading fluently at the word- and passage-level (e.g., Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, & Shanahan, 2001). Second, these programs are conducted at critical times during child development when such skills may be acquired more easily (e.g., Lyon & Chahabra, 1996).

Children identified by weak beginning reading skills who then demonstrate response-response may have been initially stymied by poorly focused or inadequately conducted classroom instruction. When the instruction is strengthened (perhaps by the adoption of a more appropriate curriculum or the addition of in-class coaching), the children's performance improves. This group of students may represent "false-positives"; those who initially appear to have a disability but in fact do not (e.g., Compton, et al., 2010; McAlenney, & Coyne, 2015).

A second response category, *response-no response* (a.k.a. "fadeout effect"), describes children who outperform controls at the end of intervention, but who do not sustain this advantage at subsequent time points (Bailey et al., 2016; Barnett, 2011; Protzko, 2015). A well-known explanation of this response type is that many schools fail to build on foundation skills taught in early intervention (Bailey et al., 2016). That is, successful early intervention programs often target children in low-income communities who, after program completion, enter more poorly resourced elementary schools that offer lower quality instruction (Crosnoe & Cooper, 2010; McLoyd, 1998; Pianta, Belsky, Houts, & Morrison, 2007; Stipek, 2004). Because of this, it is argued, skills gained in the early childhood programs eventually attenuate.

No Response-No Response and No Response-Response

No response-no response describes children whose performance is no better than controls immediately following intervention nor at follow up. Common explanations for this sustained absence of response are that the intervention is of insufficient dosage (e.g., group size is too large, duration is too brief, opportunities to respond are too infrequent); or that it does not align with students' needs; or that it fails to help them generalize learning from instructional setting to the classroom setting; or that it does not provide necessary behavioral supports (cf. L. Fuchs, Fuchs, & Malone, 2017; Wanzek & Vaughn, 2007; Wanzek & Vaughn, 2008). No response-no response children may have disabilities and may require the intensive and individualized intervention that special education programs should provide.

Finally, *no response-response* (a.k.a. "sleeper effect") describes children whose performance shows no immediate benefit from intervention but who outperform controls at a later point (Barnett, 2011; Holmes et al., 2009). Two explanations of this seemingly improbable response category are (a) skills targeted for training are difficult to acquire and (b) training effects are measured on outcomes that require transfer. Put differently, it may require the post-intervention development of children to see intervention effects on outcome measures.

Holmes et al. (2009), for example, explained that their working memory training improved the mathematics performance of at-risk children at delayed follow-up despite an absence of improvement immediately following their intervention because they needed time to apply their strengthened working memory to mathematics learning. Because this "sleeper effect" has been observed (albeit infrequently) in cognitive training (Holmes et al., 2009) and early intervention research (Barnett, 2011), we include no response-response in this taxonomy.

To be clear, the taxonomy is presented as heuristic. It is meant to illustrate a more complex way of thinking about response. To explore the utility of this perspective (not the just-

described taxonomy), we returned to a previously reported efficacy study of a reading program we developed for at-risk first graders on whom we had collected pretreatment, posttreatment, and 1-year follow-up data (cf. D. Fuchs et al., 2019). In addition to exploring multiple response types, we attempted to determine whether various child characteristics predict membership in the response groups. In recent years, research teams have searched for child characteristics that predict response to reading instruction. Most have focused on domain-specific or domain-general skills. Few have focused on both.

Predictors of Response

Domain-Specific

Several research groups have found domain-specific skills (e.g., decoding) predictive of intervention response (Clemens, Oslund, Kwok, Fogarty, Simmons, & Davis, 2019; Coyne, McCoach, Ware, Austin, Loftus-Rattan, & Baker, 2019; D. Fuchs et al., 2019; Nelson, Benner, & Gonzalez, 2003; Scarborough, 1998; Vaughn, Roberts, Capin, Miciak, Cho, & Fletcher, 2019). Coyne et al. (2019) and Vaughn et al. (2019) determined that stronger initial reading and language skills predicted a more positive response. However, Clemens et al. (2019) and D. Fuchs et al. (2019) obtained an opposite result: *Weaker* initial skills predicted more positive responses. Adding to this inconsistency is that other researchers have found no relationship between initial reading/language skills and response to reading intervention (e.g., Wanzek, Roberts, Vaughn, Swanson, & Sargent, 2019).

Domain-General

The predictive utility of domain-general skills has been more consistent, but only with respect to mathematics. Several research teams have reported that stronger cognitive skills, especially stronger working memory, predict more impressive response to early mathematics instruction (e.g., Barner, Alvarez, Sullivan, Brooks, Srinivasan, & Frank, 2016; L. Fuchs et al., 2014; Swanson, 2015). L. Fuchs et al. (2014) and Swanson (2015) obtained aptitude-by-treatment interactions such that at-risk children with relatively strong working memory benefited more from interventions involving strategy- and fluency-building, whereas at-risk children with weaker working memory profited from interventions targeting conceptual understanding.

The importance of cognitive skills to response to reading intervention is less clear. In Stuebing, Barth, Molfese, Weiss, and Fletcher's (2009) meta-analysis of 22 studies, relatively low correlations ($r = .08 \sim .27$) were obtained for pretreatment IQ and posttreatment reading performance. Stuebing, Barth, Trahan, Reddy, Miciak, and Fletcher's (2015) more recent synthesis similarly suggested the cognitive skills (attention, nonverbal reasoning, and working memory) of elementary-age at-risk students only weakly explained response ($r = .15 \sim .31$). So, whereas both of these reviews suggest cognitive skills only modestly predict response to reading instruction, this conclusion should be viewed in light of several methodological considerations. First, much of Stuebing et al.'s (2009) analysis was based on estimation because of missing data. Second, in Stuebing et al. (2015), there were few effect sizes for each cognitive skill, which may have under-powered analyses. Third, in

both reviews, the importance of domain-general skills was not examined with respect to the long-term outcomes of reading interventions. Such exploration is infrequent in the literature.

Present Study

In the current study, we used a latent profile model based on word reading performance at pretreatment, posttreatment, and 1-year follow-up (with the control group serving as a local norm at each time point) to pursue two related objectives. First, to determine if there were multiple response types in a treatment group of 265 at-risk first-grade readers who generally benefitted from an empirically-validated reading program (D. Fuchs et al., 2019). Second, whether pretreatment domain-general skills (working memory, non-verbal reasoning) and domain-specific skills (letter knowledge, decoding, passage comprehension, and language) differentiated among the types of response. Our hope was that such inquiry would shed greater light on how individual differences affect response to generally beneficial beginning reading programs; and how we might improve the prediction of risk status and individual response to such instruction and facilitate the eventual development of early interventions tailored to specific children. Last, we emphasize the exploratory nature of this work. Our intent is *not* to offer a new response taxonomy. Rather it is to encourage practitioners, policymakers, and researchers to reflect more seriously on the nature of response to instruction.

Methods

Participants

Participants were 265 children identified by their classroom teachers as “at-risk readers” in fall of first-grade. They came from 24 elementary schools in Nashville and were part of a greater number of at-risk first-graders who had been recruited for a multi-year large-scale RCT of an intensive reading program (D. Fuchs et al., 2019). All of the children were tested on multiple reading measures (see Screening Measures below). A factor score based on these measures was derived for each child and the children were rank-ordered by this score. Those at or above the 50th percentile were eliminated from study participation. So, too, were those performing below a T-score of 37 (percentile rank of 10) on the *Vocabulary* and *Matrix Reasoning* subtests of the *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999) who, we believed, were likely to have intellectual disabilities. The remaining youngsters, including the 265 in the current study, were then assigned randomly to two treatment conditions and a control group. Few of these children participated in supplemental reading interventions offered by their schools.

The RCT was implemented in four successive years. Children in the current study had been assigned to treatment groups in Years 2, 3, and 4 of the multi-year project. Hence, they represented three cohorts. Each cohort was followed into second grade. Students from Year 1 were not included in this study because the Year 1 intervention was different from the intervention in the subsequent years. Table 1 and Table 2 present study measures, child demographics and test performance, and correlations among variables. Across cohorts, standard scores for treated students on the *Sight Word Efficiency* subtest of the *TOWRE* (Torgesen, Wagner, & Rashotte, 1999) were 98 (45th percentile), 106, (66th percentile), and

98 (45th percentile) at pre-treatment, post-treatment, and follow-up testing. Before beginning the study, we received all necessary approvals and permissions from our university, the school district, and the teachers and parents/guardians of participating students.

Intervention

As indicated, there were two treatment conditions. The first focused on strengthening decoding and reading fluency; the second, on improving decoding, reading fluency, and reading comprehension. Both treatments were delivered one-to-one, three times per week, 45 min per session, for 21 weeks, totaling 47.25 hours of instruction (45 min per session x 63 sessions). Fidelity of treatment implementation was measured by direct observation at multiple time points and was found to be consistently strong (> 90%). See D. Fuchs et al. (2019) for a description of tutor training and documentation of treatment fidelity.

Students in the two treatment groups together performed significantly more strongly than controls on word reading and reading comprehension (Hedges $g > .37$, $ps < .01$), but did not perform differently from each other ($ps > .15$). On this basis, the treatment groups were combined for analyses in this paper. (See D. Fuchs et al., 2019, for a detailed description of treatment conditions and study findings).

Screening Measures

We used *The Rapid Letter Naming Test* (RLN; D. Fuchs, Fuchs, Thompson, Al Otaiba, Yen...O'Connor, 2001) as one of two letter knowledge tasks. The RLN measures the number of letters named correctly in 60 seconds. *The Rapid Sound Naming Test* (RLS; D. Fuchs et al., 2001) measures the number of letter sounds named correctly in 60 seconds. Scores are adjusted if students finish in less than 60 seconds. Test-retest reliability for the RLN is .92 among first-grade students (D. Fuchs et al., 2001).

We relied on three measures of word reading. The Sight Word Efficiency subtest of the TOWRE (Torgesen et al., 1999) measures the number of sight words identified correctly in 45 seconds. It consists of 104 sight words arranged from easiest to most difficult. The manual reports an alternate-form coefficient of .97 for a sample of 6 year olds. *Word Identification Fluency A* (WIF A; L. Fuchs, Fuchs, & Compton, 2004) comprises 2, 50-word lists. Words on each list were randomly selected from 133 high-frequency words of the Dolch pre-primer, primer, and first-grade-level lists. The 50 words were printed randomly on a page; students have 60 sec to read each list; and their score is the mean number of words read correctly across them.

If a student requires less than 60 sec, the score is pro-rated. The range of scores in fall of first grade in a representative local sample was 0 to 67, with a mean of 27 ($SD = 15$). Alternate-form reliability is .95 to .97 at first grade (L. Fuchs et al., 2004). For the *Word Identification* subtest of the *Woodcock Reading Mastery Test-Revised* (WRMT-R; Woodcock, 1998), students read sight words arranged from easiest to most difficult. The manual reports a .98 split-half reliability coefficient for a first-grade sample.

There were two measures of non-word reading. The *Phonemic Decoding Efficiency* subtest of the TOWRE (Torgesen et al., 1999) requires students to read pseudo-words (e.g., pim)

presented from easiest to most difficult. Alternate-form reliability for 6-years-olds is .97. For the *Word Attack* subtest of the WMRT-R (Woodcock, 1998), students also read pseudo-words arranged from easiest to most difficult. Split-half reliability for first-graders is .94.

IQ was assessed by the Vocabulary and Matrix Reasoning subtests of the WASI (Wechsler, 1999). For the Vocabulary subtest, children identify pictures and define words. For Matrix Reasoning, one selects 1 of 5 options that best completes a visual pattern. Test-retest reliabilities for children between 6 and 11 are reported as .85 and .76, respectively, for Vocabulary and Matrix Reasoning.

Domain-Specific Measures

To index the first-grade sample's decoding skills, we used TOWRE-Phonemic Decoding Efficiency (Torgesen et al., 1999) and WMRT-R-Word Attack (Woodcock, 1998). The Word Attack subtest of the WMRT-R consists of 45 pseudo-words, arranged from easiest to most difficult. Testing is discontinued when the student incorrectly answers 6 consecutive items. The manual reports split-half reliability for a first-grade sample as .94.

Letter knowledge was assessed by The Rapid Letter Naming Test (D. Fuchs et al., 2001), already described. To measure language, we used the Vocabulary subtest of the WASI (Wechsler, 1999), also already described, and the *Woodcock-Johnson III, Oral Comprehension* subtest (WJ-III; Woodcock, McGrew, & Mather, 2001), for which a child listens to short sentences or short passages and provides a missing word. The score is the number of items answered correctly. Test-retest reliability has been reported as .80 for 6- and 7-year-olds (Woodcock et al., 2001). Reading comprehension was measured by the *Passage Comprehension* subtest of the WRMT-R (Woodcock, 1998). Children are required to point to the picture corresponding to a rebus, or to a printed word, and then silently read passages to identify missing words. The score is the number of items answered correctly. Split-half reliability at first grade is .94 (Woodcock, 1998).

Domain-General Measures

Working memory was measured with the *Listening Recall* from the *Working Memory Test Battery for Children* (Pickering & Gathercole, 2001). The child listens to a series of short sentences, judges the veracity of each by responding "yes" or "no," and then recalls the final word of each of the sentences in sequence. There are six trials at each set size (1 to 6 sentences per set). The score is the number of trials recalled correctly. Cronbach's alpha was .85 for the current sample. We modified test administration by giving feedback on the first three items, thereby lowering the test's floor. We discontinued testing when three items in a set were answered incorrectly. The Matrix Reasoning subtest of the WASI (Wechsler, 1999) was used to assess non-verbal reasoning. Cronbach's alpha was .75 for the current sample.

Procedures

Research Assistants (RAs)—Across the multi-year large-scale implementation study, 12-16 RAs were selected each year in a competitive process that yielded full-time graduate students in education policy, special education, and teaching and learning who had

experience working with young children. They worked 20-hour weeks on the implementation study and were rigorously trained to tutor at-risk readers in the two treatment groups and to test them and control children at pre- and posttreatment in first grade. A separate and smaller group of RAs were also recruited to test the first-grade sample in their second-grade year (see below).

Fall and Spring Testing in First Grade—RAs individually tested children in their schools in four testing sessions, each of which lasted 60 min. The RAs were unfamiliar to the children and unaware of study purposes or hypotheses. Prior to fall testing, two project staff trained the RAs on multiple occasions. Each training session began with staff explaining the purpose and content of the tests and then modeling their proper administration. The RAs next role-played as examiner and examinee and obtained corrective feedback from staff.

Following training, but still before pretreatment testing, the RAs were required to practice test administration with a partner for 10 hours and then to “test” project staff on all measures. Project staff recorded their performances on a checklist. The RAs were required to achieve at least 90% accuracy when administering and scoring every test. If they performed less well, they completed additional training and tried again to meet administration and scoring criteria. No testing of children was permitted before they did so.

RA training in spring of first grade was similar to that in fall, with the exception that spring training was abbreviated because of the RAs’ familiarity with the tests. Tests were again individually administered to children in four 60-min sessions. All sessions in fall and spring were audiotaped. Twenty percent of test sessions in fall and spring were randomly selected in equal proportions across testers. Administration fidelity and scoring accuracy were measured with detailed checklists. The fidelity and accuracy scores exceeded 90% at both time points. The scores were double checked by another RA and inter-rater agreement was 90% or better.

Spring Testing in Second Grade—As mentioned, first-grade study participants were also tested in spring of second grade. RAs were recruited and trained by staff in the same manner as RAs were recruited and trained to assess the children’s first-grade performance. And the same first-grade procedures were applied at second-grade to assure testing fidelity and scoring accuracy, which were comparable to the earlier first-grade estimates.

Data Analysis

We used Little's MCAR test (Little & Rubin, 2014) on demographic, predictor, and outcome variables, which indicated that missing data from our data collection effort were missing completely at random, $\chi^2 = 21.75$, $df = 31$, $p = .89$. A majority of these missing data (i.e., five or more data points for a given variable) came from posttreatment word reading (14 data points), follow-up word reading (29 data points), and IEP status (14 data points). We compared children with missing data on these three variables to the student sample on other variables and did not find any significant differences, $p > .05$. Thus, we used Full Information Maximum Likelihood estimate for the latent class models. Because participants were nested within schools, we adjusted standard errors to account for the nested structure

using TYPE = COMPLEX with maximum likelihood with a robust estimator. We reiterate that, for the current study, our reading outcome was word reading; specifically, performance on the Sight Word Efficiency subtest of the TOWRE (Torgesen et al., 1999)

Multiple Response Types—To address our first research question, latent profile analysis was conducted with the word reading score across three time points. We used the mean and standard deviation of the control group at each time point to transform treated students' scores into *z* scores. We did so because we viewed our controls as a useful local norm; a more appropriate counterfactual, or benchmark, than the normative population of the TOWRE to understand the treated children's word-reading development. (See Lemons, Fuchs, Gilbert, & Fuchs [2014] on the importance of local controls as counterfactuals in longitudinal research.) A *z* score of 0 indicated performance that was comparable to controls; a positive *z* score and negative *z* score signaled performance that was stronger and weaker than controls, respectively. *Z* scores, in short, showed how much improvement the treated students made beyond what was expected of them absent the treatment. Because we were modeling time-series data and were not able to model curvature with the three time points, we allowed word reading at each point to co-vary within class in the model using *Mplus* 8.1 (Muthén & Muthén, 1998-2015).

We wish to clarify here that latent profile analysis is a model-based technique for identifying an unobserved categorical variable that divides a population into mutually exclusive and exhaustive latent classes. When conducting such modeling, one typically begins with an estimate of a two-profile model. Then, the number of profiles in each subsequent model increases by one until there is no longer improvement in model fit. We used the following fit statistics for comparing the latent profile models: (a) (sample-size) Adjusted Bayesian Information Criterion (ABIC), (b) Lo-Mendel-Rubin Likelihood Ratio Test (LMR-LRT), (c) entropy, and (d) posterior probability. Due to the modeling of nested data using TYPE = Complex option, a bootstrapped likelihood ratio test was not possible. We chose ABIC over Bayesian Information Criterion because ABIC is better for relatively small samples such as ours (Nylund, Asparouhov, & Muthén, 2007).

Smaller values are preferred for ABIC. Differences in ABIC greater than 10 indicate decisive differences in model fit (Kass & Raftery, 1995). The LMR-LRT is a direct test of the model with *k* profiles against a model with *k*-1 profiles. A significant LMR-LRT indicates the model with *k* profiles fits the data significantly better than a model with *k*-1 profiles. Model entropy (ranging from 0 to 1) is an index of how well each case's most likely profile membership corresponds to the actual profile membership in the model (i.e., model classification certainty). A higher value indicates a better model fit (Jung & Wickrama, 2008; Hart, Logan, Thompson, Kovas, McLoughlin, & Petrill, 2016). For each group profile, an average posterior probability value was computed. An average probability of .90 indicates 90% of individuals for whom a profile is most likely are actually placed in that profile by the analysis. Following the recommendations of Nylund et al. (2007), we considered all fit statistics when determining a best model.

Individual Differences and Response Types—Our second research question was whether there were pretreatment differences that might predict membership in one or more

of the response-profile groups. We first created composite scores using principal component analysis (Bartholomew, Steele, Galbraith, & Moustaki, 2008) for letter knowledge (rapid sound naming and rapid letter naming), decoding (phonemic decoding efficiency and word attack), and language (vocabulary and listening comprehension). Then, predictors of response-profile membership were added to the latent analyses as auxiliary variables. We examined the extent to which each domain-specific and domain-general skill predicted class membership. Predictors were evaluated by the R3STEP function, which involves multinomial logistic regression (i.e., regressing class membership on all predictors simultaneously). To reduce the type I error, we applied Benjamini-Hochberg correction procedure for each predictor on all possible response-profile group comparisons.

Results

Table 1 presents descriptive statistics. Overall, the first-grade sample's performance was normally distributed. Table 2 shows correlations among all variables. Table 3 displays fit statistics for the latent profile models. It indicates ABIC decreased until the four-profile model. Entropy was largest for the five-profile model. LMR-LRT was not significant for any model. The four- and five-profiles both had relatively better fit statistics than the other profile models. A closer look at sample size for the response groups in the five-profile model revealed a single-student group. Thus, we chose the four-profile model for subsequent analyses. Average latent class probabilities for membership in Profiles 1-4, respectively, were .89 .90, .89, and .85.

Table 4 shows that members of Profile 1 ($n = 45$; 17% of treated children) had averaged word-reading z scores of .25, 1.64, and 1.26 at pretreatment (grade 1), posttreatment (grade 1), and follow-up (grade 2), respectively. Averaged z scores of those in Profile 2 ($n = 109$; 41% of treated children) were .30, .47, and .55 at the three time points. Profile 3 ($n = 90$; 34% of treated children) mean z scores were $-.11$, $-.15$, and $-.55$. And for Profile 4 members ($n = 21$; 8% of the treated children), mean z scores were -1.24 , -1.26 , and -1.57 . Percentile scores based on the normative population of the TOWRE Sight Word Efficiency subtest were as follows: for Profile 1 children, 48, 84, and 79 at pre- and posttreatment and follow-up, respectively; for Profile 2 students, 48, 70, and 55; for the Profile 3 group, 42, 55, and 27; and for Profile 4 children, 32, 29, and 8 (see numbers in parentheses in Table 4).

Comparisons among the four response groups revealed that students in Profiles 1 and 2 showed stronger word reading at all time points than children in Profiles 3 and 4. Although Profile 1 and Profile 2 children read comparably at pretreatment, Profile 1 children read better at posttreatment and at 1-year follow-up. In comparison to Profile 4 children, Profile 3 members read better at all three time points. Based on the z scores and group comparisons, we describe the four profile groups—1 through 4, respectively— as “strongly responsive” children with stronger initial performance; “mildly responsive” children with stronger initial performance; “mildly non-responsive” children with stronger initial performance; “strongly non-responsive” children with weaker initial performance.

We plotted the developmental trajectories of every treated student in each profile group at pre- and posttreatment and follow-up. The mean trajectories of the four groups are displayed

in Figure 1. Trajectories of the individual members of each group, in addition to the groups' mean trajectories, are shown in color in Supplemental Figure S1, which can be found on line.

Next, we ran analyses on pretreatment domain-specific and domain-general skills to explore individual differences that might predict, or characterize, membership in the four response groups. As Table 5 shows, among domain specific skills, letter knowledge and passage comprehension differentiated the groups. In particular, the mildly responsive group showed stronger pretreatment letter knowledge than the mildly non-responsive group and greater pretreatment letter knowledge and passage comprehension than the strongly non-responsive group. The strongly responsive group showed more impressive pretreatment passage comprehension than the strongly non-responsive group. The mildly non-responsive group showed better pretreatment passage comprehension than the strongly non-responsive group. We did not find pretreatment domain-general skills (i.e., working memory and non-verbal reasoning) to significantly predict the four response types.

Discussion

We focused on the responses of 265 weak first-grade readers to a generally efficacious reading program (cf. Fuchs et al., 2019). We applied a latent profile model to the children's reading performance prior to and immediately following program completion, and again in spring of second grade. The primary purpose of the analysis was to determine if the children's responses could legitimately be characterized in terms of response/no response—the conventional conception of “response” in RTI frameworks—or whether they reflected a greater number of response types. We also explored whether fall-of-first-grade domain-specific skills and domain-general skills predicted membership in the response categories.

Typology of Response

Our analysis yielded four, not two, response types: two *responsive* groups (“strongly” and “mildly”) and two *non-responsive* groups (“strongly” and “mildly”). As mentioned, Figure 1 shows the mean trajectories of the four groups across pre- and posttreatment and follow-up. Supplemental Figure S1 (on line) displays the trajectories of every child in the four (color-coded) groups, and the (bolded) mean group trajectories. Whereas there is notable variation within the response categories, and considerable overlap between them, they are clearly distinguishable.

Strongly Responsive and Mildly Responsive—Discussion of the strongly responsive and mildly responsive groups cannot be separated from the nature of our reading program. As described, it emphasized decoding, reading fluency, and (in one of two treatment groups) reading comprehension. It was implemented with fidelity and intensity: one-to-one, three times per week, 45 min per session, for 21 weeks (47.25 hours). Thus, it should not be surprising that 58% (154 of 265) of the treated children (41% mildly responsive; 17% strongly responsive) were superior to controls on word reading at posttreatment in grade 1 and at follow-up in grade 2 (z scores $> .50$; Table 4).

The mildly responsive and strongly responsive groups were comparable to each other and to controls at pretreatment (z scores = .25 ~.30; see Table 4). However, at posttreatment and follow-up, the two treated groups demonstrated stronger reading performance than controls (z scores = .47 ~ 1.64), and adequate-or-better performance with respect to the TOWRE Sight Word Efficiency norms (55th ~ 84th percentile; see scores in parentheses in Table 4). Assuming the gains of the children in the two treated groups are attributable to treatment, our findings jibe with Torgesen et al.'s (1999) claim that early intervention can reduce the number of students at risk for serious reading problems (as well as false positives). That said, we do not regard the responsive children's performance (compared to controls and TOWRE norms) as an indication that our program transformed them from "at-risk" to "normalized" readers. More persuasive evidence of such a transformation would have required the addition of a typically-developing group to our study design.

Strongly Non-Responsive and Mildly Non-Responsive—We also found 111 of 265 treated children (42%) who were either strongly non-responsive (8%) or mildly non-responsive (34%) to the reading program. In contrast to the strongly non-responsive group, whose performance was inferior to that of controls at all time points (z scores = -1.57 ~ -1.24), the mildly non-responsive group performed similarly to controls at pre- and posttreatment in first grade (z scores = $-.15$ ~ $-.11$), but lagged behind in second grade when they were no longer program participants (z score = $-.55$). May we interpret the mildly non-responsive group's second-grade performance as indirect evidence of program effects? We lack confidence in such a conclusion because of necessary inference-making, but it is possible that our program kept this non-responsive group abreast of average-performing controls in first grade when they were being tutored and, when the tutoring was withdrawn, they could no longer keep up.

If such an interpretation is right, one might expect this group of weak readers to become more responsive to more intensive forms of our standard program—perhaps if it were delivered 5 days per week instead of 3 days, or conducted for 60 min per session instead of 45 min—and more responsive to an instructional protocol with greater focus on word reading than reading comprehension. In other words, our data, if correctly interpreted here, suggest we might involve these students in a modified, more intensive version of the standard version of our program.

Not so for the strongly non-responsive group. These most vulnerable children, we believe, would require instruction tailored to their specific needs, which, by definition, is not something a standard instructional approach provides. In the last 40 years, researchers have developed a version of tailored instruction known as data-based instruction (a.k.a. "systematic formative evaluation" and "curriculum-based assessment"), and have validated it for use with students with very significant learning problems (see Deno & Mirkin, 1977; L. Fuchs & Fuchs, 1986; L. Fuchs, Fuchs, Hamlett, & Stecker, 2020; Stecker, Fuchs, & Fuchs, 2005).

The larger point here is that differentiating among non-responsive students may lead to the identification of one or more subgroups that require qualitatively different interventions to make progress. In this sense, our use of latent profile analysis to identify subgroups of non-

responders, and to begin to describe them by domain-specific and domain-general characteristics, is a step (albeit a modest step) toward developing aptitude x treatment interactions (ATIs; e.g., Cronbach, 1957; 1975). In principle, ATIs represent a two-step research strategy: Step 1 requires precise and empirically-validated description of the learner characteristics of a targeted population; Step 2 requires development of interventions that explicitly address these characteristics. Steps 1 and 2, in short, require correlational and experimental research, respectively. Of course, our (Step 1) descriptive findings on strongly and mildly non-responsive subgroups require validation.

Response Profiles We Did Not Find—We did not find a “no response-response” profile. Previous research suggests that it may occur when skills targeted for training are difficult to improve and outcome measures require the transfer of taught skills to different skills or different contexts. In such circumstances, time may be necessary for children to master the more difficult-to-acquire skills and apply them in novel situations (e.g., Barnett, 2011; Holmes et al., 2009). Because early reading skills are relatively teachable (e.g., National Reading Panel, 2000), our intervention was intensively delivered, and our word-reading measures were aligned with the treatment, it may be understandable that we did not find evidence of unresponsiveness in the short term and responsiveness later.

However, for severely delayed readers, a pattern of “no response-response” might be possible if the children are continuously engaged in quality, appropriately-focused intervention. In a meta-analysis by Wanzek and Vaughn (2007), treatment dosage seemed important. Heavy-dose reading programs (100 sessions or more) produced considerably more positive outcomes than programs conducted less intensively.

We also failed to find a profile of “response-no response” in our sample, which may reflect the likelihood that decoding, fluency, and comprehension skills are developed outside or beyond the intervention program. That is, children who initially responded to the reading program may have used acquired skills to further develop their reading competence (e.g., Stanovich, 2009), which is not easily weakened by subsequent misguided instruction.

Predictors of Variation in Response

We did not find pretreatment predictors that differentiated the two (mildly and strongly) responsive groups from each other. However, the mildly responsive group was superior to the mildly non-responsive group in letter knowledge and superior to the strongly non-responsive group in both decoding and passage comprehension. The strongly responsive group showed better pretreatment passage comprehension than the strongly non-responsive group; and the mildly non-responsive group showed better pretreatment decoding and passage comprehension than the strongly non-responsive group.

There may be at least two meaningful conclusions based on our prediction effort. First, domain-specific skills differentiated (a) the responsive groups from the non-responsive groups and (b) the two non-responsive groups from each other, highlighting the importance of domain-specific skills in early reading development (e.g., Chall, 1983; Ehri, 2005; National Reading Panel, 2000; Stanovich, 2009). Reading comprehension differentiated the two responsive groups from the strongly non-responsive group and the two non-responsive

groups from each other (see Table 5). This finding persisted even when word-level skills (letter knowledge and decoding) and language skills were considered simultaneously in the analyses.

In other words, the young at-risk students with *relatively* strong comprehension were more likely to improve their word reading performance, and maintain it one year later, after intensive intervention. Those with weaker pretreatment reading comprehension were less likely responsive as indicated by their performance at posttreatment and follow-up. This suggests a bi-directional relationship between early word reading and reading comprehension development, as noted by others before us (e.g., Etmanskie, Partanen, & Siegel, 2016; Hoover & Gough, 1990).

A second conclusion is that pretreatment domain-general skills (working memory, non-verbal reasoning) did not predict the latent profiles, which is consistent with some prior reading research (e.g., Stuebing et al., 2009; 2015). It is inconsistent, however, with mathematics research, which shows that domain-general skills predict response to intervention (e.g., Barner et al., 2017; L. Fuchs et al., 2014; Swanson, 2015).

That domain-general skills only modestly predict response to reading programs is perplexing because of the considerable evidence that reading problems are related to cognitive deficits (e.g., Carretti, Borella, Cornoldi, & De Beni, 2009; Peng, Wang, Tao, Sun, 2016; Spencer & Wagner, 2018). A close look at this literature suggests that the strongest relations between cognitive processes and reading are when the two are respectively represented by working memory and reading comprehension. Because much of reading intervention research targets word-level reading and relies on word reading as an outcome, future studies might investigate whether working memory and other executive functions more strongly predict latent profile groups among children in *reading comprehension* programs whose responses to these programs are evaluated on reading comprehension measures.

Study Limitations and Future Research

There are at least several study limitations. First, the failure of domain-general processes to predict membership in the four response categories may be partly explained by the limited variance associated with our cognitive measures. Larger samples and more sensitive measures would likely strengthen exploration of the predictive utility of these processes. Second, we did not collect classroom-level information (e.g., the nature of instruction) as study participants advanced from first to second grade and moved between classes and schools. Future studies might look at whether classroom and school factors influence who is and is not responsive.

Third, the exclusion from the sample of children with very low vocabulary no doubt influences the generalizability of our findings. Language difficulty predicts reading difficulties and, therefore, the at-risk readers whom we included may not allow for a comprehensive understanding of service delivery systems like RTI for a whole-school population.

Finally, our findings are inextricably connected to our methods—e.g., we involved young at-risk readers in a very particular intervention and, for purposes of these analyses, we only used word reading to index their response. Whereas they are consistent with a more nuanced understanding of instructional response, they are the product of a study that must be taken for what it is: an exploration. However provocative they may be, they are suggestive, and suggestive only. They may not be seen as validating a new response taxonomy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by grants R01HD056109, P20HD075443, and HD15052 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD). The authors are responsible for the paper's content, which does not necessarily represent the views of the NICHD or National Institutes of Health. We thank Brett Miller (NICHD) for his advice and encouragement; the teachers and administrators of the Metropolitan Nashville Public Schools for their interest and cooperation; and David Francis for his suggestion that we display the developmental trajectories of every treated student in the study sample.

References

- Bailey D, Duncan GJ, Odgers CL, & Yu W (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, 10(1), 7–39. doi:10.1080/19345747.2016.1232459 [PubMed: 29371909]
- Bailey DH, Nguyen T, Jenkins JM, Domina T, Clements DH, & Sarama JS (2016). Fadeout in an early mathematics intervention: Constraining content or preexisting differences? *Developmental Psychology*, 52(9), 1457–1469. doi: 10.1037/dev0000188 [PubMed: 27505700]
- Barnett WS (2011). Effectiveness of early educational intervention. *Science*, 333(6045), 975–978. doi:10.1126/science.1204534 [PubMed: 21852490]
- Barner D, Alvarez G, Sullivan J, Brooks N, Srinivasan M, & Frank MC (2016). Learning mathematics in a visuospatial format: A randomized, controlled trial of mental abacus instruction. *Child Development*, 87(4), 1146–1158. doi: 10.1111/cdev.12515 [PubMed: 27062391]
- Barth AE, Stuebing KK, Anthony JL, Denton CA, Mathes PG, Fletcher JM, & Francis DJ (2008). Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences*, 18(3), 296–307. doi:10.1016/j.lindif.2008.04.004 [PubMed: 19081758]
- Bartholomew DJ, Steele F, Moustaki I, & Galbraith JI (2008). *Analysis of multivariate social science data* (2nd ed.). Boca Raton, FL: CRC. doi:10.1080/10705511.2011.607725
- Carretti B, Borella E, Cornoldi C, & De Beni R (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, 19(2), 246–251. doi: 10.1016/j.lindif.2008.10.002
- Chall JS (1983). *Learning to read: The great debate*. New York: McGraw-Hill.
- Clemens NH, Oslund E, Kwok OM, Fogarty M, Simmons D, & Davis JL (2019). Skill moderators of the effects of a reading comprehension intervention. *Exceptional Children*, 85(2), 197–211. doi: 10.1177/0014402918787339
- Compton DL, Fuchs D, Fuchs LS, Bouton B, Gilbert JK, Barquero LA, ... & Crouch RC (2010). Selecting 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(2), 327–340. doi: 10.1037/a0018448 [PubMed: 20689725]
- Coyne MD, McCoach DB, Ware S, Austin CR, Loftus-Rattan SM, & Baker DL (2019). Racing against the vocabulary gap: Matthew effects in early vocabulary instruction and intervention. *Exceptional Children*, 85(2), 163–179. doi: 10.1177/0014402918789162

- Cronbach LJ (1957). The two disciplines of scientific psychology. *American Psychologist*, 12 (11), 671–684. doi: 10.1037/h0043943
- Cronbach LJ (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30(2), 116–127. doi: 10.1037/h0076829
- Crosnoe R, & Cooper CE (2010). Economically disadvantaged children’s transitions into elementary school: Linking family processes, school contexts, and educational policy. *American Educational Research Journal*, 47(2), 258–291. doi:10.3102/0002831209351564 [PubMed: 20711417]
- Deno SL, & Mirkin PK (1977). *Data-based program modification: A manual* University of Minnesota: Minneapolis. ERIC-ED 144270
- Ehrhardt J, Huntington N, Molino J, & Barbaresi W (2013). Special education and later academic achievement. *Journal of Developmental & Behavioral Pediatrics*, 34(2), 111–119. doi: 10.1097/DBP.0b013e31827df53f [PubMed: 23363974]
- Ehri LC (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167–188. doi: 10.1207/s1532799xssr0902_4
- Ehri LC, Nunes SR, Willows DM, Schuster BV, Yaghoub-Zadeh Z, & Shanahan T (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel’s meta-analysis. *Reading Research Quarterly*, 36(3), 250–287. doi: 10.1598/RRQ.36.3.2
- Etmanskie JM, Partanen M, & Siegel LS (2016). A longitudinal examination of the persistence of late emerging reading disabilities. *Journal of Learning Disabilities*, 49(1), 21–35. doi: 10.1177/0022219414522706 [PubMed: 24596111]
- Fletcher JM, Lyon GR, Fuchs LS, & Barnes MA (2019). *Learning disabilities: From identification to intervention* (2nd ed.). New York, NY: Guilford.
- Fletcher JM, & Miciak J (2019). *The identification of specific learning disabilities: A summary of research on best practices*. Texas Center for Learning Disabilities.
- Foorman BR, Francis DJ, Shaywitz SE, Shaywitz BA, & Fletcher JM (1997). The case for early reading intervention In Blachman BA (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 243–264). Mahwah, NJ: Erlbaum.
- Fuchs D, & Fuchs LS (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41(1), 93–99. doi: 10.1598/RRQ.41.1.4
- Fuchs D, Fuchs LS, & Compton DL (2004). Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disability Quarterly*, 27(4), 216–227. doi: 10.2307/1593674
- Fuchs D, Fuchs LS, Compton DC, Bryant J, & Davis GN (2008). Making “secondary intervention” work in a three-tier responsiveness-to-intervention model: Findings from the first-grade longitudinal reading study at the National Research Center on Learning Disabilities. *Reading and Writing: An Interdisciplinary Journal*, 21(4), 413–436. doi: 10.1007/s11145-007-9083-9
- Fuchs D, Fuchs LS, Thompson A, Otaiba SA, Yen L, Yang NJ, ... & O'Connor RE (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology*, 93(2), 251–267. doi: 10.1037/0022-0663.93.2.251
- Fuchs D, Kearns DM, Fuchs LS, Elleman AM, Gilbert JK, Patton S, ... & Compton DL (2019). Using moderator analysis to identify the first-grade children who benefit more and less from a reading comprehension program: A step toward Aptitude-by-Treatment Interaction. *Exceptional Children*, 85(2), 229–247. doi: 10.1177/0014402918802801 [PubMed: 30636789]
- Fuchs LS, & Fuchs D (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53, 199–208. doi: 10.1177/001440298605300301 [PubMed: 3792417]
- Fuchs LS, Fuchs D, & Compton DL (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children*, 71(1), 7–21. doi: 10.1177/001440290407100101
- Fuchs LS, Fuchs D, Hamlett C, & Stecker PM (2020). Bringing data-based individualization to scale: A call for the next-generation technology of teacher supports. Submitted.
- Fuchs LS, Fuchs D, & Malone A (2017). The taxonomy of intervention intensity. *Teaching Exceptional Children*, 50(1), 38–43. ED571640. doi: 10.1177/0040059917703962

- Fuchs LS, Schumacher RF, Sterba SK, Long J, Namkung J, Malone A, ... & Changas P (2014). Does working memory moderate the effects of fraction intervention? An aptitude–treatment interaction. *Journal of Educational Psychology*, 106(2), 499–514. doi: 10.1037/a0034341
- Gilmour AF, Fuchs D, & Wehby JH (2019). Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities. *Exceptional Children*, 85, 329–346. doi: 10.1177/0014402918795830
- Hart SA, Logan JA, Thompson L, Kovas Y, McLoughlin G, & Petrill SA (2016). A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *Journal of Educational Psychology*, 108(2), 181–193. doi: 10.1037/edu0000045 [PubMed: 26957650]
- Holmes J, Gathercole SE, & Dunning DL (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, 12(4), 9–15. doi: 10.1111/j.1467-7687.2009.00848.X
- Hoover WA, & Gough PB (1990). The simple view of reading. *Reading and Writing*, 2(2), 127–160. doi: 10.1007/BF00401799
- Jung T, & Wickrama KAS (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2(1), 302–317. doi: 10.1111/j.1751-9004.2007.00054.x
- Kass RE, & Raftery AE (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. doi: 10.2307/2291091
- Lemons CJ, Fuchs D, Gilbert JK, & Fuchs LS (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher*, 43(5), 242–252. doi: 10.3102/0013189X14539189
- Little RJ, & Rubin DB (2014). *Statistical analysis with missing data*. John Wiley & Sons.
- Lyon GR, & Chhabra V (1996). The current state of science and the future of specific reading disability. *Mental Retardation and Developmental Disabilities Research Reviews*, 2(1), 2–9. doi: 10.1002/(SICI)1098-2779(1996)2:1<2::AID-MRDD2>3.0.CO;2-X
- McAlenney AL, & Coyne MD (2015). Addressing false positives in early reading assessment using intervention response data. *Learning Disability Quarterly*, 35(1), 53–65. doi: 10.1177/0731948713514057
- McLoyd VC (1998). Socioeconomic disadvantage and child development. *American Psychologist*, 53(2), 185–204. doi: 10.1037/0003-066X.53.2.185 [PubMed: 9491747]
- Miller B, McCardle P, & Hernandez R (2010). Advances and remaining challenges in adult literacy research. *Journal of Learning Disabilities*, 43(2), 101–107. doi: 10.1177/0022219409359341 [PubMed: 20179305]
- Muthén LK, & Muthén BO (1998–2015). *MPlus Users Guide*. Los Angeles: Author.
- National Reading Panel (US), National Institute of Child Health, & Human Development (US). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development, National Institutes of Health.
- Nelson RJ, Benner GJ, & Gonzalez J (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice*, 18(4), 255–267. doi: 10.1111/1540-5826.00080
- North Carolina Department of Public Instruction, Exceptional Children Division. (2017, 3 24). *Evaluation and identification of specific learning disabilities*. Author.
- Nylund KL, Asparouhov T, & Muthén BO (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535–569. doi: 10.1080/10705510701575396
- Peng P, Wang C, Tao S, & Sun C (2017). The deficit profiles of Chinese children with reading difficulties: A meta-analysis. *Educational Psychology Review*, 29(3), 513–564. doi: 10.1007/s10648-016-9366-2
- Pianta RC, Belsky J, Houts R, & Morrison F (2007). Opportunities to learn in America's elementary classrooms. *Science*, 315(5820), 1795–1796. doi: 10.1126/science.1139719 [PubMed: 17395814]
- Pickering S, & Gathercole SE (2001). *Working Memory Test Battery for Children (WMTB-C)*. New York: Psychological Corporation.

- Protzko J (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence*, 53, 202–210. doi: 10.1016/j.intell.2015.10.006
- Scarborough HS (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*, 48(1), 115–136. doi: 10.1007/s11881-998-0006-5
- Spencer M, & Wagner RK (2018). The comprehension problems of children with poor reading comprehension despite adequate decoding: A meta-analysis. *Review of Educational Research*, 88(3), 366–400. doi: 10.3102/0034654317749187 [PubMed: 29785063]
- Stanovich KE (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education*, 189(1–2), 23–55. doi: 10.1177/0022057409189001-204
- Stein DJ (2012). Dimensional or categorical: Different classifications and measures of anxiety and depression. *Medicographia*, 34(3), 270–275.
- Stecker PM, Fuchs L, & Fuchs D (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42(8), 795–819. doi: 10.1002/pits.20113
- Stipek D (2004). Teaching practices in kindergarten and first grade: Different strokes for different folks. *Early Childhood Research Quarterly*, 19(4), 548–568. doi: 10.1016/j.ecresq.2004.10.010
- Stuebing KK, Barth AE, Molfese PJ, Weiss B, & Fletcher JM (2009). IQ is not strongly related to response to reading instruction: A meta-analytic interpretation. *Exceptional Children*, 76(1), 31–51. doi: 10.1177/001440290907600102 [PubMed: 20224749]
- Stuebing KK, Barth AE, Trahan LH, Reddy RR, Miciak J, & Fletcher JM (2015). Are child cognitive characteristics strong predictors of responses to intervention? A meta-analysis. *Review of Educational Research*, 85(3), 395–429. doi: 10.3102/0034654314555996 [PubMed: 26535015]
- Swanson HL, Harris KR, & Graham S (Eds.). (2013). *Handbook of learning disabilities*. New York: Guilford.
- Swanson HL (2015). Cognitive strategy interventions improve word problem solving and working memory in children with math disabilities. *Frontiers in Psychology*, 6, 1099. doi: 10.3389/fpsyg.2015.01099 [PubMed: 26300803]
- Torgesen JK, Alexander AW, Wagner RK, Rashotte CA, Voeller KK, & 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(1), 33–58. doi: 10.1177/002221940103400104 [PubMed: 15497271]
- Torgesen JK, Wagner RK, & Rashotte CA (1999). *TOWRE: Test of Word Reading Efficiency*. Austin TX: PRO-ED.
- Tran L, Sanchez T, Arellano B, & Lee Swanson H (2011). A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities*, 44(3), 283–295. doi: 10.1177/0022219410378447 [PubMed: 21521870]
- Vandell DL, Belsky J, Burchinal M, Steinberg L, Vandergrift N, & NICHD Early Child Care Research Network. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD study of early child care and youth development. *Child Development*, 81(3), 737–756. doi: 10.1111/j.1467-8624.2010.01431.x [PubMed: 20573102]
- Vaughn S, Roberts G, Capin P, Miciak J, Cho E, & Fletcher JM (2019). How initial word reading and language skills affect reading comprehension outcomes for students with reading difficulties. *Exceptional Children*, 85(2), 180–196. doi: 10.1177/0014402918782618 [PubMed: 31223172]
- Vaughn S, Wanzek J, Wexler J, Bath A, Cirino PT, Fletcher J, Denton C Roberts G & Francis D (2010). The relative effects of group size on reading progress of older students with reading difficulties. *Reading and Writing*, 23(8), 931–956. doi: 10.1007/s11145-009-9183-9 [PubMed: 21072131]
- Wanzek J, Roberts G, Vaughn S, Swanson E, & Sargent K (2019). Examining the role of pre-instruction academic performance within a text-based approach to improving student content knowledge and understanding. *Exceptional Children*, 85(2), 212–228. doi: 10.1177/0014402918783187

- Wanzek J, & Vaughn S (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities*, 41(2), 126–142. doi: 10.1177/0022219407313426 [PubMed: 18354933]
- Wanzek J, & Vaughn S (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, 36(4), 541–561.
- Wanzek J, & Vaughn S (2009). Students demonstrating persistent low response to reading intervention: Three case studies. *Learning Disabilities Research & Practice*, 24(3), 151–163. doi: 10.1111/j.1540-5826.2009.00289.x
- Wechsler D (1999). Wechsler Abbreviated Scale of Intelligence. Psychological Corporation
- Woodcock RW (1998). Woodcock Reading Mastery Tests-Revised. Circle Pines, MN: American Guidance Service.
- Woodcock RW, McGrew KS, & Mather N (2001). Woodcock-Johnson III Tests of Cognitive Abilities. Itasca, IL: Riverside Pub.

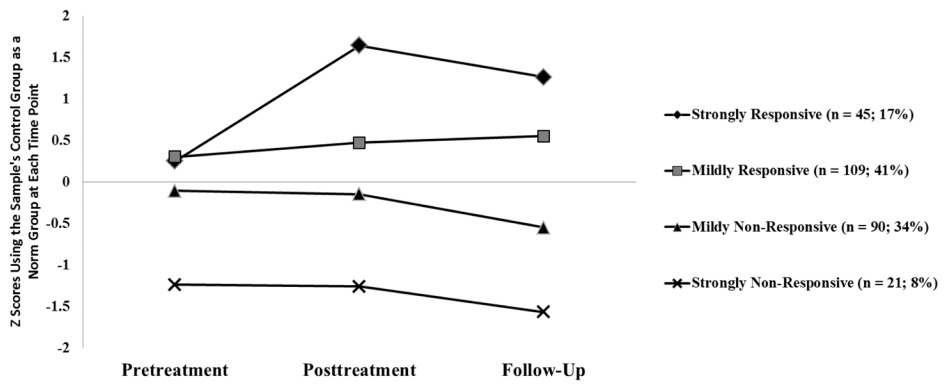


Figure 1.
Latent Profile Groups across Three Time Points

Table 1

Demographics and Test Performance

Variables	N	Mean	SD	Skewness	Kurtosis	Min	Max
Gender (male)	264 (147)	-	-	-	-	-	-
FRL (yes)	262 (220)	-	-	-	-	-	-
IEP (yes)	251 (20)	-	-	-	-	-	-
Race (Caucasian/Black/Hispanic)	264 (77/117/43)	-	-	-	-	-	-
ELL (yes)	261 (34)	-	-	-	-	-	-
Rapid Sound Naming	265	26.99	11.02	0.33	1.59	0	76
Rapid Letter Naming	265	37.40	12.41	-0.15	0.45	0	71
Phonemic Decoding Efficiency	265	2.52	2.82	1.09	0.48	0	12
Word Attack	265	2.78	2.97	0.87	-0.39	0	11
Vocabulary	265	16.63	6.20	-0.24	-0.03	0	33
Listening Comprehension	265	12.13	4.77	-0.17	-0.53	0	22
Passage Comprehension	265	8.87	2.83	0.13	-0.53	3	17
Non-verbal Reasoning	265	6.69	4.00	1.14	1.57	0	22
Listening Recall	265	2.22	3.08	1.20	0.67	0	12
TOWRE sight word pretreatment*	265	10.03 (.03)	4.73	-0.23	-0.34	0	22
TOWRE sight word posttreatment ⁺	251	32.22 (.31)	10.11	0.09	-0.08	7	56
TOWRE sight word follow-up ⁺	236	43.81 (.14)	11.99	-0.32	-0.46	12	70

Note. N = the number of data points. FRL: Free/Reduced lunch status; ELL: English language learning status; Rapid Sound Naming: Rapid Sound Naming Test (D. Fuchs et al., 2001); Rapid Letter Naming: Rapid Letter Naming Test (D. Fuchs et al., 2001); Phonemic Decoding Efficiency: TOWRE-Phonemic Decoding Efficiency (Torgesen et al., 1999); Word Attack: WMRT-R-Word Attack (Woodcock, 1998); Vocabulary: WASI-Vocabulary (Wechsler, 1999); Listening Comprehension: WJ3-Oral Comprehension Subtest (Woodcock, 2001); Passage Comprehension: WRMT-R-Passage Comprehension Subtest (Woodcock, 1998); Non-verbal Reasoning: WASI-Matrix Reasoning Subtest (Wechsler, 1999); Listening Recall: WMTB-Listening Recall (Pickering & Gathercole, 2001); TOWRE sight word pretest, posttest, and follow-up: The Sight Word Efficiency subtest of the TOWRE (Torgesen et al., 1999) at the beginning of 1st grade, the end of 1st grade, and the end of 2nd grade, respectively.

Z scores of TOWRE performance at three time points are in parentheses and are based on the control group's performance.

Table 2

Correlations among Variables

Pre-treatment Measures	1	2	3	4	5	6	7	8	9	10	11	12
1. Rapid Sound Naming	1.00											
2. Rapid Letter Naming	.42**	1.00										
3. Phonemic Decoding Efficiency	.34**	.27**	1.00									
4. Word Attack	.31**	.18**	.67**	1.00								
5. Vocabulary	.18**	.13*	.21**	.15*	1.00							
6. Listening Comprehension	.19**	.13*	.21**	.17**	.53**	1.00						
7. Passage Comprehension	.15*	.22**	.31**	.33**	.15*	.25**	1.00					
8. Non-verbal Reasoning	0	.07	.16**	.25**	.13*	.31**	.28**	1.00				
9. Listening Recall	.23**	.18**	.25**	.31**	.30**	.37**	.31**	.27**	1.00			
10. TOWRE sight word pretreatment	.30**	.39**	.34**	.34**	.14*	.21**	.63**	.23**	.16**	1.00		
11. TOWRE sight word posttreatment	.27**	.26**	.23**	.28**	.03	.16**	.46**	.21**	.17**	.47**	1.00	
12. TOWRE sight word follow-up	.22**	.24**	.24**	.22**	.02	.09	.31**	.14*	0.11	.34**	.81**	1.00

Note:

* $p < .05$

** $p < .001$.

Rapid Sound Naming; Rapid Sound Naming Test (D. Fuchs et al., 2001). Rapid Letter Naming; Rapid Letter Naming Test (D. Fuchs et al., 2001). Phonemic Decoding Efficiency; TOWRE-Phonemic Decoding Efficiency (Torgesen et al., 1999). Word Attack; WMRT-R-Word Attack (Woodcock, 1998). Vocabulary; WASI-Vocabulary (Wechsler, 1999). Listening Comprehension; WJ 3-Oral Comprehension Subtest (Woodcock, 2001). Passage Comprehension; WRMT-R-Passage Comprehension Subtest (Woodcock, 1998). Non-verbal Reasoning; WASI-Matrix Reasoning Subtest (Wechsler, 1999). Listening Recall; WMTB-Listening Recall (Pickering & Gathercole, 2001). TOWRE Sight Word Efficiency pretreatment, posttreatment, and follow-up scores at the beginning of 1st grade, end of 1st grade, and end of 2nd grade, respectively.

Table 3

Model Fit Statistics from Latent Profile Analyses on the Treatment Group

Model for The Treatment Group	ABIC	ABIC	Entropy	LMR-LRT
2 Profiles	4980.64		.79	25.46, $p = .23$
3 Profiles	4970.75	-9.89	.67	18.69, $p = .51$
4 Profiles	4960.64	-10.11	.79	18.91, $p = .10$
5 Profiles	4961.70	1.06	.82	8.21, $p = .29$
6 Profiles	4961.89	.19	.72	9.04, $p = .80$

Note. ABIC: Sample Size Adjusted Bayesian Information Criterion. LMR-LRT = Lo-Mendel-Rubin Likelihood Ratio Test. ABIC: Changes on ABIC are between the current set of profiles and previous set of profiles.

Performance on TOWRE Sight Word Reading for the Four Profile Groups at Pre- and Posttreatment and Follow-Up

Table 4

Time	Profile 1 (Strongly Responsive, n = 45)			Profile 2 (Mildly Responsive, n = 109)			Profile 3 (Mildly Non-Responsive, n = 90)			Profile 4 (Strongly Non-Responsive, n = 21)		
	Estimated Mean (%tile)	SE	z	Estimated Mean (%tile)	SE	z	Estimated Mean (%tile)	SE	z	Estimated Mean (%tile)	SE	z
Pretreatment	11.09 (48)	.63	0.25	11.32 (48)	.53	0.30	9.40 (42)	.59	-0.11	4.07 (32)	.61	-1.24
Posttreatment	47.25 (84)	.67	1.64	33.95 (70)	.68	0.47	26.95 (55)	.77	-0.15	14.31 (29)	.84	-1.26
Follow-Up	58.52 (79)	.78	1.26	49.22 (55)	.81	0.55	34.73 (27)	.79	-0.55	21.26 (8)	1.63	-1.57

Note. TOWRE Sight Word reading performance at fall (pretreatment) and spring (posttreatment) of 1st grade and spring (follow-up) of 2nd grade. In parentheses, percentile rankings (%tile) were based on the standardization population of the TOWRE. Z scores were calculated with controls as a local norm at each time point.

Table 5
 Pretreatment Domain-General and Domain-Specific Skills Predicting Membership in the Profile Groups

Variables	Strongly Responsive vs. Mildly Non-Responsive		Strongly Responsive vs. Mildly Non-Responsive		Strongly Responsive vs. Mildly Non-Responsive		Strongly Responsive vs. Mildly Non-Responsive		Strongly Responsive vs. Mildly Non-Responsive	
	Coef (SE)	p	Coef (SE)	p	Coef (SE)	p	Coef (SE)	p	Coef (SE)	p
Working Memory	.08	.17	-.02	.83	-.14	.48	-.10	.26	-.22	.53
Non-Verbal Reasoning	-.004	.94	.06	.25	.16	.12	.06	.15	.16	.29
Letter Knowledge	-.28	.43	.36	.36	2.13	.02	.64	.02	2.41	.01
Decoding	-.50	.18	-.10	.70	-.13	.77	.40	.12	.37	.94
Language	.12	.71	-.19	.60	-.34	.45	-.31	.15	-.46	.79
Passage Comprehension	.12	.25	.28	.003	.73	< .001	.15	.08	.61	.006

Note. Coef = Coefficient; SE = Standard Error.

Letter Knowledge: The factor score of Rapid Sound Naming (D. Fuchs et al., 2001) and Rapid Letter Naming (D. Fuchs et al., 2001). Decoding: The factor score of TOWRE-Phonemic Decoding Efficiency (Torgesen et al., 1999) and WMRT-R-Word Attack (Woodcock, 1998). Language: The factor score of WASI-Vocabulary (Wechsler, 1999) and WI 3-Oral Comprehension Subtest (Woodcock, 2001). Passage Comprehension: WRMT-R-Passage Comprehension Subtest (Woodcock, 1998). Non-verbal Reasoning: WASI-Matrix Reasoning Subtest (Wechsler, 1999). Working Memory: WMTB-Listening Recall (Pickering & Gathercole, 2001).

Bolded coefficients are significant after Benjamini-Hochberg correction.