

# The Construct and Predictive Validity of a Dynamic Assessment of Young Children Learning to Read: Implications for RTI Frameworks

Douglas Fuchs<sup>1</sup>, Donald L. Compton<sup>1</sup>, Lynn S. Fuchs<sup>1</sup>,  
Bobette Bouton<sup>1</sup>, and Erin Caffrey<sup>2</sup>

## Abstract

The purpose of this study was to examine the construct and predictive validity of a dynamic assessment (DA) of decoding learning. Students ( $N = 318$ ) were assessed in the fall of first grade on an array of instruments that were given in hopes of forecasting responsiveness to reading instruction. These instruments included DA as well as one-point-in-time (static) measures of early alphabetic knowledge, rapid automatized naming (RAN), phonemic awareness, oral vocabulary, listening comprehension, attentive behavior, and hyperactive or impulsive behavior. An IQ test was administered in spring of second grade. Measures of reading outcomes administered in spring of first grade were accuracy and fluency of word identification skills and reading comprehension. Factor analysis using principal axis factor extraction indicated that DA loaded on a first factor that also included language abilities and IQ, which the authors refer to as the “language, IQ, and DA” factor. It was relatively distinct from two additional factors: (a) “speeded alphabetic knowledge and RAN” and (b) “task-oriented behavior.” A three-level (children nested within classroom; classrooms nested within school) random intercept model with fixed effects predictors suggested that DA differed from word attack in predicting future reading skill and that DA was a significant predictor of responsiveness to instruction, contributing unique variance to end-of-first-grade word identification and reading comprehension beyond that explained by other well-established predictors of reading development.

## Keywords

RTI, dynamic assessment, beginning reading assessment

Many teachers, administrators, policy makers, and professional organizations are looking to responsiveness to intervention (RTI) as an educational reform that will provide early intervention to at-risk learners and promote more valid identification of children with learning disabilities. RTI is viewed by many as a more valid identification process than traditional psychometric approaches because it guarantees in principle that all children will participate in scientifically validated, generally effective instruction in a multilevel system of service delivery. Thus, the use of RTI is expected to accelerate the academic achievement of most children and reduce the likelihood that the unttaught or poorly taught will be misidentified as disabled.

## RTI: Instruction as Test

More specifically, RTI is seen as a more valid method of disability identification because the scientifically validated, generally effective instruction becomes a “test”—as much a

test as the *Wide Range Achievement Test* or *Stanford-Binet*. Instruction is the test stimulus and the student’s level or rate of performance is her response. “Responsive” students to Tier 1 (or “core”) instruction continue with it; responsive students to Tier 2 instruction reenter Tier 1 instruction. Unresponsive students in Tier 1 move to Tier 2, unresponsive students in Tier 2 move to Tier 3, and so on. In short, responsiveness to instruction determines the level of instructional intensity that is necessary for a given child; and responsiveness, or its absence, can help determine whether the child has a disability.

<sup>1</sup>Vanderbilt University, Nashville, TN, USA

<sup>2</sup>Congressional Research Service, Washington, DC, USA

## Corresponding Author:

Douglas Fuchs, Vanderbilt University, Peabody #228,  
230 Appleton Place, Nashville, TN 37203-5721  
Email: [doug.fuchs@vanderbilt.edu](mailto:doug.fuchs@vanderbilt.edu)

Just as commercial publishers and professional groups properly concern themselves about the validity of scores from test instruments (see American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), practitioners using RTI need to attend to the validity of their instruction. Choosing research-principled curricula and empirically validated academic programs that address at-risk students' learning and implementing the curricula and programs with fidelity are necessary to ensure the meaningfulness of "instruction as test." If practitioners choose invalid or unvalidated curricula and programs or implement validated versions of them without fidelity and if the instruction is not generally effective, it becomes impossible to interpret a child's unresponsiveness to such instruction.

As reasonable as it may sound, instruction-as-test raises legitimate issues. Among the most important is that instruction at each tier in an RTI framework may last 8 to 10 weeks or more. Thus, a chronically unresponsive child may fail across multiple tiers for 30 weeks before practitioners recognize that she or he requires most intensive instruction (e.g., special education). This scenario prompts the question, "Can we conceive of an identification process that more quickly identifies nonresponders in need of much greater help?" Adding to the importance of this question is the sense that many practitioners can identify such children with considerable accuracy. The question then becomes, "How do we do this as precisely as possible?" Dynamic assessment (DA), although unconventional and infrequently used by practitioners, might be a valuable component of a modified and more flexible and useful RTI process (e.g., Grigorenko, 2009). Below, we briefly describe DA by comparing it to conventional, or traditional, assessment.

### *A More Efficient RTI?*

**DA and traditional tests.** A widely known criticism of the use of traditional one-point-in-time tests with low-achieving children is that many of them do not accurately predict future academic performance. This inaccuracy is partly because of "floor effects." That is, many unskilled kindergartners and first graders obtain a score of zero when administered a traditional reading test such as the Word Identification subtest of the *Woodcock Reading Mastery Tests*. The zero score can reflect at least two important limitations. First, many traditional tests do a poor job of sampling (i.e., devoting too few items to) basic or elementary skills. Second, they typically assess only two states: unaided success or unaided failure. From a Vygotskian perspective, however, children may be somewhere between these two states: unable to perform the task independently but able to achieve success with assistance. With DA, the examiner can explore the amount and nature of this assistance. Thus, DA is an index of a child's readiness to change and as such

represents a unique means of differentiating performance among children at the low end of the achievement continuum (e.g., Spector, 1992).

DA has been described as the assessment of learning potential, mediated learning, testing the limits, mediated assessment, and assisted learning and transfer by graduated prompts. Across these various conceptions, DA differs from traditional testing in terms of the nature of the examiner–student relationship, content of feedback, and emphasis on process rather than product (Grigorenko & Sternberg, 1998).

In traditional testing, the examiner is a neutral participant who provides standardized directions but not, typically, performance-contingent feedback. Many DA examiners, by contrast, not only give performance-contingent feedback but also offer instruction in response to student failure to alter or enhance the student's performance. Put differently, traditional testing is oriented toward the product (i.e., level of performance) of student learning, whereas the DA examiner's interest is in both the product and the process (i.e., rate of growth) of the learning. Some claim this twin focus on the level and rate of learning makes DA a better predictor of future performance. It may help decrease the number of "false positives," or children who seem at risk but who, with timely instruction, may respond relatively quickly and perform within acceptable limits. As mentioned, data from DA may also help identify the type and intensity of intervention necessary for academic success. It incorporates a test–teach–test format, conceptually similar to RTI techniques. However, it can potentially measure one's responsiveness within a much shorter time frame.

Notwithstanding the logic and promise of such an approach, there has been infrequent research of DA's psychometric properties (see Grigorenko & Sternberg, 1998), especially regarding its predictive validity. Caffrey, Fuchs, and Fuchs (2008) conducted a review of the research exploring DA's predictive validity and found that traditional tests and DA similarly predicted future academic performance, irrespective of whether the students were normally achieving or at risk. However, DA seemed to tap achievement differently than traditional achievement tests and cognitive tests and therefore may be an important supplement in efforts to identify responders and nonresponders to instruction.

**Purpose of this article.** Several years ago we began to develop a DA measure of early reading (i.e., decoding). The Caffrey et al. (2008) review indicated a need to explore both its construct and predictive validity, construct validity because DA appears to measure achievement differently than do traditional achievement and cognitive tests, predictive validity because a DA instrument with strong predictive validity may help identify students at risk for school failure who need more intensive intervention. That is, DA in conjunction with traditional testing may indicate more accurately a student's potential for change and likelihood of school success as well as appropriate instruction for the

student. Hence, our interest in DA should not be seen as support for the exclusion of traditional tests from prediction making, nor are we suggesting DA as a substitute for multiple tiers of increasingly intensive instruction. Rather, it might add accuracy and efficiency as part of a test battery that helps practitioners match children in a timely manner to appropriate tiers of instruction within an RTI framework.

## Method

### Participants

**Students.** Students were selected from 56 first-grade classrooms in seven Title I and seven non-Title I schools in urban and suburban middle Tennessee. We assessed every consented child ( $N = 712$ ) on word identification fluency, rapid letter naming, and rapid sound naming, which are described below. Latent class analysis facilitated a sorting of these 712 first graders into high-, average-, and low-performing groups. We then randomly sampled the members of each group, deliberately oversampling low-performing children—the group most likely to be targeted for DA—to increase their numbers in our prediction models.

Of an initial sample of 485 children, 310 were low performing, 83 were average performing, and 92 were high performing. All were assessed in fall of first grade with a prediction battery and in spring with an outcome battery. In addition, we assessed IQ in the spring of second grade. Of the initial 485 children, complete data were available for 416 after administration of the prediction battery, 378 after the outcome battery, and 320 after we collected IQ data in spring of second grade. Also, 2 children were eliminated as “outliers” during data analysis, leaving us with a final sample of 318. Among this group, 50.6% were female and 44.1% were African American, 41.0% were Caucasian, 6.0% were Hispanic, 2.4% were Asian, and 6.5% were “other.” Of the students, 42% received free or reduced-price lunch. There were no reliable differences between the initial sample ( $n = 485$ ) and the group that completed the outcome battery in spring of first grade ( $n = 378$ ) on rapid letter naming,  $F(1, 414) = 0.046, p = .830$ , rapid sound naming,  $F(1, 414) = 0.122, p = .727$ , word identification fluency (WIF),  $F(1, 414) = 0.553, p = .458$ , gender,  $\chi^2(1) = 0.789, p = .674$ , race,  $\chi^2(6) = 1.12, p = .976$ , or free or reduced-price lunch status,  $\chi^2(1) = 0.230, p = .632$ . However, the initial sample performed less well than the final sample ( $n = 318$ ) on rapid sound naming,  $F(1, 414) = 5.06, p = .025$ , and WIF,  $F(1, 414) = 4.59, p = .033$ , but not on rapid letter naming,  $F(1, 414) = 3.74, p = .054$ . There were no between-group differences in terms of gender,  $\chi^2(1) = 2.79, p = .248$ , race,  $\chi^2(6) = 4.26, p = .640$ , or free or reduced-price lunch status,  $\chi^2(1) = 0.07, p = .933$ .

**Examiners.** There were nine examiners, all of whom were master’s students in various academic departments at

Peabody College of Vanderbilt University. The project coordinator and a doctoral student trained the examiners in test administration in nine meetings for a total of 10 hr. In addition, the examiners practiced administering the tests to each other and were required to administer each test to the project coordinator with a minimum of 90% accuracy. Examiners were given detailed guidelines, describing “desirable” test locations in their school buildings, how to communicate with the students and to build rapport, and what to do in case of unforeseen school events such as fire drills.

### Procedure

The first-grade study children were tested in 2006–2007. The prediction battery consisted of the following variables (and measures): alphabetic knowledge (rapid letter naming, rapid sound naming, and word attack), rapid automatized naming (RAN; letters and digits), phonemic awareness (elision and sound matching), oral vocabulary, DA, listening comprehension, and teacher ratings of students’ attentive behavior and hyperactivity or impulsivity. As indicated, in the spring of second grade, an IQ test was added to the prediction battery. (The timing of the IQ testing, although not ideally synchronized with the administration of the other measures, seems acceptable given its stability.) The outcome battery, administered in spring of first grade, comprised an untimed measure of word identification, a timed measure of sight word reading, and a test of reading comprehension. Following is more specific description of the tests used to select participants and the prediction and outcome batteries. Well-known tests are described briefly.

### Measures Used to Initially Select Participants

With WIF (L. S. Fuchs, Fuchs, & Compton, 2004), children are given a single page of 50 high-frequency words randomly sampled from the Dolch preprimer, primer, and first-grade-level lists. They have 1 min to read the words as quickly as they can. If they hesitate on an item for 4 s, the tester prompts them to proceed to the next word. Split-half reliability for the sample was .95. For rapid letter naming (D. Fuchs et al., 2001), the tester presents a page with 52 letters (all 26 letters in uppercase and lowercase) displayed in random order. Students have 1 min to say the letter names. The score is the number of correct letters. Rapid sound naming (D. Fuchs et al., 2001) requires the tester to present a page with 26 lowercase letters displayed in random order. Students have 1 min to say sounds. The score is the number of correct sounds. For rapid letter naming and rapid sound naming, test-retest reliability exceeds .94. For all three measures, if the child finishes before 1 min, the score is prorated.

## Prediction Battery

**Alphabetic knowledge, RAN, and phonemic awareness.** We used three measures of alphabetic knowledge. Rapid letter naming (D. Fuchs et al., 2001) and rapid sound naming (D. Fuchs et al., 2001) were just described. The *Woodcock Reading Mastery Test–R/NU* Word Attack (WA; Woodcock, 1998) requires children to pronounce pseudowords presented in list form. Our two measures of RAN were rapid digit naming and rapid letter naming from the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner, Torgesen, & Rashotte, 1999). With two more CTOPP measures, we explored the predictive value of phonemic awareness. For elision (Wagner et al., 1999), the tester says words, which the child reproduces with a syllable or phoneme removed; for sound matching (Wagner et al., 1999), the tester presents a word and the student determines which of three words (presented as pictures) start with the same sound.

**Language-related measures.** Our oral vocabulary measure was taken from the *Woodcock-Johnson Psychoeducational Battery–Revised* (Woodcock, McGrew, & Mather, 2001). The tester says words, for which the child provides synonyms or antonyms. A listening comprehension measure exploring students' understanding of sentences and passages came from the *Woodcock Diagnostic Reading Battery* (Woodcock, 1998). We also administered the two-subtest *Wechsler Abbreviated Scale of Intelligence* (Psychological Corporation, 1999), which yields an IQ score linked to the third edition of the *Wechsler Intelligence Scale for Children*. The first of these subtests, Vocabulary, comprises 42 items that measure expressive vocabulary, verbal knowledge, and foundation of information. Matrix Reasoning, the second subtest, measures nonverbal fluid reasoning with 35 items. Testers present a series of patterns; students select "missing pieces" from five choices.

**Attentive behavior and hyperactivity or impulsivity.** The SWAN is an 18-item teacher rating scale (www.adhd.net). Items reflect the American Psychiatric Association's (1994) *Diagnostic and Statistical Manual of Mental Disorders* criteria for attention-deficit/hyperactivity disorder. The SWAN comprises two scales, one for inattentive behavior (Items 1–9) and a second for hyperactivity or impulsivity (Items 10–18). We report data for each (7-point) subscale as the average rating per item across the nine relevant items. Coefficient alpha for this sample exceeded .96 for the two subscales.

**Dynamic assessment.** We created a one-session DA (D. Fuchs et al., 2007). The tester uses a scripted standardized administration and pseudowords to teach three increasingly difficult decoding skills: CVC (taught as linguistic word families), CVCe, and CVC(C)ing. For each skill, five levels of more and more explicit scaffolded instruction can be used. Between each level of instruction, six pseudowords (not used for instruction but paralleling instructional items) are presented. If the student reads at least five of the six

words correctly, the tester deems the skill mastered and moves the student to the next, more advanced skill. If the student reads fewer than five words correctly, the tester engages the student in more explicit instruction of that skill. If the student fails to achieve mastery across Levels 1–5 of the scaffolded instruction for the given skill, the DA session is terminated. Put differently, if the student does not reach mastery by Level 5 of the CVC skill, the CVCe and doubling sections of the test are not administered. If the student does not reach mastery on the CVCe skill, the doubling section of the test is not administered. (Contact the first author for the script that directed the administration of the CVC section of our DA.)

For each of the three decoding skills, or sections of the test, students are scored 1 through 5. A score of 1 indicates the student reached mastery after the first level of instruction (Level 1); a score of 5 indicates that the student reached mastery at the fifth and final level (Level 5). If students are not administered a section (because of a lack of mastery of a more elementary skill), they are given a score of 5. Scores for the three skills or sections of the DA are added for a total score. A lower score indicates quicker mastery of content. In our analyses, we reversed values so that higher scores indicated quicker mastery. In a pilot study with 100 first-grade children, the 4-week stability coefficient for the DA score was .72 (Fuchs, 2009).

## Outcome Battery

**Word identification.** We included two measures of word identification, the first of which was the *Woodcock Reading Mastery Test–R/NU* Word Identification (WID; Woodcock, 1998). Children read words presented in list form. We also administered Sight Word Reading Efficiency (SWE), a subtest of the *Test of Word Reading Efficiency* (Torgesen, Wagner, & Rashotte, 1997). It allows children 45 s to read words presented in list form.

**Reading comprehension.** The first set of items on the *Woodcock Reading Mastery Test–R/NU* Passage Comprehension (PC; Woodcock, 1998) requires the tester to present a symbol (i.e., rebus) and to ask the child to point to the picture corresponding to the rebus. Next, the child points must point to the picture representing words printed on the page. Finally, the child reads a passage silently and identifies the missing word.

## Results

### Construct Validity

To explore the construct validity of DA, we conducted an exploratory factor analysis on the first-grade prediction battery, which included the second-grade IQ measure. Factor analysis is meant to reveal latent variables that cause



**Table 1.** Distributional Characteristics of the Manifest Variables in the Factor Analysis

Measures	<i>M</i>	<i>SD</i>	Skew	Kurtosis
Rapid letter naming <sup>a</sup>	44.0	15.9	.003	-.026
Rapid sound naming <sup>a</sup>	30.3	11.9	.000	-.014
Rapid automatized naming—Letters <sup>b</sup>	9.7	3.1	.005	-.025
Rapid automatized naming—Numbers <sup>b</sup>	9.5	2.8	.002	-.013
Elision <sup>b</sup>	10.1	2.9	-.009	-.028
Sound matching <sup>b</sup>	9.9	2.3	-.029	.102
Oral vocabulary <sup>c</sup>	96.8	14.2	.000	-.019
Dynamic assessment <sup>d</sup>	9.7	3.6	-.128	-.399
Listening comprehension <sup>c</sup>	95.6	16.5	.005	-.061
Attentive behavior <sup>d</sup>	37.4	12.8	-.009	-.074
Hyperactive or impulsive behavior <sup>d</sup>	37.7	13.1	-.030	-.169
IQ <sup>c</sup>	98.5	14.9	.001	-.013

Note: *N* = 318.

<sup>a</sup>Items per second.

<sup>b</sup>Scaled scores (*M* = 10, *SD* = 3).

<sup>c</sup>Standard scores (*M* = 100, *SD* = 15).

<sup>d</sup>Raw scores.

performance to covary on the “manifest” variables (our measures). During factor extraction, the shared variance of a variable is partitioned from its unique variance and error variance to reveal the underlying factor structure. Only shared variance appears in the solution. Manifest variables were measures of alphabetic knowledge (rapid letter naming and rapid sound naming), RAN (letters and digits), phonemic awareness (elision and sound matching), oral vocabulary, DA, listening comprehension, IQ, and teacher ratings of attentive behavior and hyperactive or impulsive behavior.

We first explored whether underlying assumptions of factor analysis were met in our data set. As mentioned, we identified two participant outliers with *z* scores larger than 3.0 standard deviations (*SDs*) from the mean on both RAN measures (letters and digits). They were omitted from all analyses, leaving 318 children with complete data. This sample size is considered adequate for factor analysis (Tabachnick & Fidell, 2007). Table 1 presents means and *SDs* as well as skew and kurtosis estimates for the manifest variables. Distributional estimates of skew and kurtosis were not statistically significant, indicating the assumption of univariate normality was met. Pairwise scatterplots revealed strong linear relations among variables and suggested multivariate normality. Table 2 displays correlations among the manifest variables, which ranged from .17 to .76. The pattern of these correlations does not suggest the presence of multicollinearity. Overall, the data structure appeared appropriate for factor analysis.

Factor analysis using principal axis factor extraction produced three factors with eigenvalues greater than 1.0 (i.e., Kaiser–Guttman rule), which accounted for 60.2% of the variance. Examination of the scree plot confirmed a natural break point in the curve after three factors. Thus, we used a three-factor model to explain the data. The magnitude of the KMO measure of sampling (.875) and Bartlett’s test of sphericity,  $\chi^2(66) = 1983.5$ ,  $p < .001$ , indicated that factor analysis had a high probability of identifying the latent structure of the data. Commonality values were relatively high, ranging from .40 to .87. This indicated that significant variance was explained by the factor structure. An oblique rotation that allowed factors to correlate was used to examine factor structure. Table 3 displays factor loading patterns, communalities, eigenvalues, percentage variance, and factor correlations for the factor extraction. Factor loadings less than .40 were eliminated from the table.

The three factors identified are relatively clear and easy to interpret. Factor 1 is defined by high loadings on language ability (oral vocabulary, listening comprehension, elision, and sound matching), IQ, and DA. We refer to this factor as “language, IQ, and DA.” Factor 2 is defined by high loadings on rapid letter naming, rapid sound naming, and the two RAN measures. We name this factor “speeded alphabetic knowledge and RAN.” Factor 3 is defined by loadings on the attentive behavior and hyperactivity or impulsivity ratings, and we refer to it as “task-oriented behavior.” Correlations among the factors ranged from .38 to .54. There was little cross-loading of variables among them. So results indicate that DA loads on a factor that also includes language abilities (including comprehension and phonemic awareness) and IQ. This factor is relatively distinct from factors of speeded reading and task-oriented behavior.

### Predictive Validity

To assess predictive validity, we examined the extent to which DA predicted variance in the outcome battery administered in spring of first grade, after controlling for student performance on the other predictors. Before exploring the utility of the prediction models, we examined correlations and looked for potential school and classroom dependency in the outcome measures. Correlations between the predictor measures and spring-of-first-grade outcome measures are shown in Table 4. (Because high scores on the DA signify poor performance—the opposite of the scales of the traditional measures—correlations are negative.) The measures associated with the language, IQ, and DA factor and the speeded alphabetic knowledge and RAN factor were highly correlated with outcome measures, whereas measures associated with the task-oriented behavior factor were somewhat lower. Correlations between DA and reading outcomes were high, ranging from .61 to

**Table 2.** Correlations for the Manifest Variables in the Factor Analysis

Measures	1	2	3	4	5	6	7	8	9	10	11	12
1. Rapid letter naming	—											
2. Rapid sound naming	.57	—										
3. Rapid automatized naming—Letters	-.55	-.46	—									
4. Rapid automatized naming—Digits	-.54	-.50	.75	—								
5. Elision	.36	.32	-.40	-.36	—							
6. Sound matching	.41	.41	-.49	-.41	.56	—						
7. Oral vocabulary	.32	.33	-.32	-.27	.57	.53	—					
8. Dynamic assessment	-.47	-.41	.40	.38	-.60	-.61	-.49	—				
9. Listening comprehension	.34	.27	-.29	-.21	.59	.44	.58	-.39	—			
10. Attentive behavior	.39	.37	-.44	-.37	.54	.52	.43	-.47	.39	—		
11. Hyperactive or impulsive behavior	.22	.20	-.22	-.17	.38	.33	.24	-.25	.27	.76	—	
12. IQ	.32	.34	-.31	-.26	.59	.50	.59	-.52	.54	.41	.25	—

Note:  $N = 318$ . All correlations are based on raw scores and are significant at the .01 level (two-tailed).

**Table 3.** Factor Loading Patterns, Communalities, Eigenvalues, Percentage Variance, and Factor Correlations for the Factor Extraction

Measures	Factor 1	Factor 2	Factor 3	Communality ( $h^2$ )
Oral vocabulary	.809			.597
IQ	.798			.586
Listening comprehension	.699			.465
Elision	.697			.613
Dynamic assessment	-.554			.520
Sound matching	.509			.549
Rapid automatized naming—Numbers		.893		.694
Rapid automatized naming—Letters		.836		.692
Rapid letter naming		-.633		.502
Rapid sound naming		-.532		.392
Attentive behavior			.909	.867
Hyperactive or impulsive behavior			.805	.740
Eigenvalue	5.61	1.49	1.16	
Percentage variance	43.49	9.27	7.38	
Correlations				
Factor 1	—			
Factor 2	-.54	—		
Factor 3	-.49	.38	—	

Note:  $N = 318$ .

.72. The correlations between WA skill and reading outcomes seemed even stronger, ranging from .74 to .81.

In terms of dependency, school level and classroom level variance estimates and intraclass correlations are shown in Table 5 for the spring of first grade outcome battery. For each reading outcome, significant dependency existed at the levels of school and classroom. The percentage of school variance ranged from 8.2 to 12.6; the percentage of classroom variance ranged from 6.1 to 10.4. In all, approximately 18% of the variance in reading outcomes was associated with school and classroom membership.

We therefore used multilevel modeling to partition variance between child and school levels (i.e., estimating both classroom and school variance). This permitted the prediction of child level variance by itself using HLM 6.06 (Raudenbush, Bryk, & Congdon, 2004) with robust standard errors. Two different models estimated the relationship between DA and reading outcomes in spring of first grade. To examine the relationship between DA and WA as predictors, Model 1 allowed only DA and WA to compete. Model 2, by contrast, estimated the predictive utility of DA in the presence of all the fall predictors.

**Table 4.** Correlations Among Fall Predictor Measures and Spring Reading Measures

Fall Predictors	WID	SWE	PC
Letter Naming	.59	.62	.53
Sound Naming	.46	.47	.43
Rapid automatized naming—Letters	-.62	-.64	-.54
Rapid automatized naming—Digits	-.55	-.61	-.48
Elision	.67	.59	.68
Sound matching	.65	.60	.61
Oral vocabulary	.53	.48	.60
Dynamic assessment	-.72	-.61	-.65
Listening comprehension	.45	.39	.55
Word attack	.81	.74	.72
Attention	.61	.58	.61
Hyperactivity	.37	.35	.39
IQ	.57	.53	.66

Note:  $N = 318$ . WID = word identification; SWE = sight word efficiency; PC = passage comprehension.

**Table 5.** Variance Estimates and Intraclass Correlations for the First Grade Reading Outcomes

Measures	Variance Estimate			ICC-School	ICC-Class
	School	Class	Individual		
WID	30.1	14.6	193.3	.126	.061
SWE	3.1	2.5	23.4	.107	.086
PC	23.8	30.3	236.4	.082	.104

Note:  $N = 318$ . ICC = intraclass correlation coefficient; WID = word identification; SWE = sight word efficiency; PC = passage comprehension.

Table 6 shows regression coefficients, significance levels, standard errors, and pseudo  $R^2$  estimates. Results of Model 1 indicate that DA and WA together explain a substantial percentage of variance for each of the three reading outcomes, ranging from 48.5% to 63.6%. The variance shared by DA and WA ranged from 27.3% to 39.4%. The DA and WA measures each predicted unique variance across the three outcome measures with WA consistently predicting greater variance. Findings from Model 2 indicate that when DA was in competition with all predictors, it was a statistically significant predictor of word identification and PC, uniquely accounting for 2.3% and 1.0% of the variance, respectively. Results suggest that DA is a significant predictor of responsiveness to reading instruction, contributing unique variance beyond that associated with other established predictors of reading development.

## Discussion

We have described an exploration of the utility of DA of first graders who were learning to read. Findings suggest it may have value as part of a test battery to identify young children with severe learning needs who require most intensive instruction in RTI frameworks. The importance of such a battery is suggested by the fact that typical RTI procedures require many children to spend weeks and sometimes the better part of a school year in insufficiently intensive instruction before they gain access to the appropriate education they need. A test battery that identifies such children can expedite a correct match between them and an appropriate instructional environment, which has always been a *raison d'être* of RTI.

In developing DA for early reading, we were looking for evidence of its construct and predictive validity, that it both (a) taps skills untapped by more established measures and (b) explains unique variance in reading performance. Findings indicate modest support for both. As explained, DA loaded on a factor with language and IQ, suggesting it is partly a language task that requires the manipulation of novel information (not unlike IQ tests). Regarding its predictive validity, one can reasonably be of two minds. On the plus side, it explains unique variance ranging from 1.0% (PC) to 2.3% (WA). On the minus side, it can be said that these proportions are small.

When weighing the importance of these alternate views, we ask the reader to keep in mind that we explored the value of DA relative to skills and abilities that prior research has identified as important predictors of word-level reading or reading comprehension. These included alphabetic knowledge represented by measures of rapid letter and sound naming (e.g., D. Fuchs et al., 2001) and WA; RAN letters and digits (e.g., Compton, 2000); phonemic awareness represented by elision and sound matching tasks (e.g., Bus & Ijzendoorn, 1999); oral vocabulary (e.g., Carroll, 1993), listening comprehension (e.g., Joshi, Williams, & Wood, 1998); and IQ (e.g., Swanson & Alexander, 1997).

We also included teacher ratings of attentive behavior and hyperactive or impulsive behavior as a predictor to compete with DA. Although less research has focused on these teacher ratings, which might be understood as estimates of task-oriented behavior, prior work provides a basis for their inclusion in prediction models (e.g., Cutting & Scarborough, 2006). These ratings require teachers to make judgments about children's ability to attend to detail, sustain attention, listen, follow directions, organize tasks, keep track of things, ignore extraneous stimuli, remember daily activities, and stay in place. They are based on direct classroom observations and seem better connected to learning to read in classroom settings than in one-to-one testing situations

**Table 6.** Fixed Effects and Pseudo  $R^2$  Estimates for the Three-Level Random Intercept Models First Grade Outcomes

Fall Predictors	Spring Reading Outcomes					
	WID		SWE		PC	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
<b>Model 1</b>						
Intercept	42.15***	0.64	40.53***		17.66***	0.24
Word attack	0.96***	0.05	1.13***		0.28***	0.02
DA	-1.19***	0.15	-0.72***		-0.43***	0.01
Pseudo $R^2$ (%)						
WA + DA	63.6		48.5		51.1	
Unique WA	18.6		17.9		12.3	
Unique DA	5.6		3.3		5.1	
<b>Model 2</b>						
Intercept	42.19***	0.48	40.62***	0.60	17.65***	0.18
Letter naming	0.12***	0.04	0.20***	0.04	0.03*	0.01
Sound naming	-0.04	0.04	-0.03	0.05	-0.01	0.01
Rapid automatized naming—Letters	-0.13***	0.04	-0.13***	0.04	-0.03*	0.01
Rapid automatized naming—Digits	-0.03	0.04	-0.12***	0.04	-0.01	0.01
Elision	0.21	0.14	0.05	0.18	0.10	0.05
Sound matching	0.21*	0.12	0.26	0.13	0.02	0.04
Oral vocabulary	0.03	0.19	0.05	0.18	0.11*	0.06
DA	-0.64***	0.18	-0.03	0.24	-0.18**	0.06
Listening comp	-0.02	0.13	-0.16	0.13	0.09*	0.04
Word attack	0.57***	0.06	0.61***	0.09	0.10***	0.02
Attention	0.13***	0.06	0.14	0.07	0.03	0.02
Hyperactivity	0.01	0.04	0.02	0.06	0.02	0.01
IQ	0.07*	0.03	0.13***	0.04	0.07***	0.01
Pseudo $R^2$ (%)						
All Predictors	74.3		68.13		67.9	
Unique DA	2.3		0.4		1.0	

Note:  $N = 318$ . All models are based on raw scores. WID = word identification; SWE = sight word efficiency; PC = passage comprehension; Coeff. = coefficient; DA = dynamic assessment; WA = word attack; comp = comprehension.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

like DA, where the tester can better control inattentive, hyperactive, and impulsive behavior. The teacher ratings of task-oriented behavior, therefore, represent worthy competitors to DA in capturing variance in response to general education reading instruction.

This is not to say we are satisfied with our DA measure. It behaves too much like some of the traditional tests of which we have been critical. That is, it seems to have a floor effect with few low-performing children in our sample demonstrating they learned the CVC(C)ing skill. Two possible solutions come to mind. The first is to substitute a less difficult task for the CVC(C)ing task. A second would be to keep the CVC(C)ing task but to give children more helpful guidance—more explicit information—earlier in the sequence of graduated prompts. However, by extending the DA measure in this way we would be increasing its length and administration time—already at 20 to 30 min. Thus, strengthening the all-around

utility of the DA—adding efficiency—requires ongoing attention to various (and sometimes competing) psychometric and practical considerations.

### Declaration of Conflicting Interests

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

### Funding

Funding for this research was provided by Grant R324G060036 from the U.S. Department of Education (Institute of Education Sciences) and Core Grant HD15052 from the National Institute of Child Health and Human Development. An early phase of this work was conducted under the auspices of the National Research Center on Learning Disabilities and supported by Grant H324U010004 from the U.S. Department of Education (Office of Special Education Programs). Statements do not reflect the position



or policy of these agencies, and no official endorsement by them should be inferred. Similarly, with respect to Erin Caffrey, views expressed in the article are hers and her coauthors. They should not be interpreted as those of the Congressional Research Service or the Library of Congress.

## References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *The standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Bus, A. G., & Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*, 403–414.
- Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *Journal of Special Education, 41*, 254–270.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York, NY: Cambridge University Press.
- Compton, D. L. (2000). Modeling the growth of decoding skills in first-grade children. *Scientific Studies of Reading, 4*, 219–258.
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277–299.
- Fuchs, D. (2009). *Dynamic assessment and RTI*. Coronado, CA: Pacific Coast Research Conference.
- Fuchs, D., Fuchs, L. S., Compton, D. L., Bouton, B., Caffrey, E., & Hill, L. (2007). Dynamic assessment as responsiveness-to-intervention: A scripted protocol to identify young at-risk readers. *Teaching Exceptional Children, 39*, 58–63.
- Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N., . . . O'Connor, R. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology, 93*, 251–267.
- Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*, 7–21.
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities, 42*, 111–132.
- Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124*, 75–111.
- Joshi, R. M., Williams, K. A., & Wood, J. R. (1998). Predicting reading comprehension from listening comprehension: Is this the answer to the IQ debate? In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 319–327). Mahwah, NJ: Lawrence Erlbaum.
- Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Harcourt Brace.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). HOM 6 for Windows [computer software]. Lincolnwood, IL: Scientific Systems International, Inc.
- Spector, J. E. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology, 84*, 353–363.
- Swanson, H. L., & Alexander, J. E. (1997). Cognitive processes as predictors of word recognition and reading comprehension in learning-disabled and skilled readers: Revisiting the specificity hypothesis. *Journal of Educational Psychology, 89*, 128–158.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.
- Woodcock, R. W. (1998). *Woodcock Reading Mastery Test—Revised/Normative Update*. Circle Pines, MN: AGS.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Psychoeducational Ability*. Itasca, IL: Riverside.

## About the Authors

**Douglas Fuchs**, PhD, is a professor of special education at Peabody College of Vanderbilt University.

**Donald L. Compton**, PhD, is an associate professor of special education at Peabody College of Vanderbilt University.

**Lynn S. Fuchs**, PhD, is a professor of special education at Peabody College of Vanderbilt University.

**Bobette Bouton** was a project coordinator at Vanderbilt and is now a doctoral student at the University of Georgia.

**Erin Caffrey**, PhD, was a research assistant and is now with the Congressional Research Service.