


Development of First-Graders' Word Reading Skills: For Whom Can Dynamic Assessment Tell Us More?

Journal of Learning Disabilities
2017, Vol. 50(1) 95–112
© Hammill Institute on Disabilities 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0022219415599343
journaloflearningdisabilities.sagepub.com


Eunsoo Cho, PhD¹, Donald L. Compton, PhD², Jennifer K. Gilbert, PhD³,
Laura M. Steacy, PhD², Alyson A. Collins, PhD⁴, and Esther R. Lindström, MEd³

Abstract

Dynamic assessment (DA) of word reading measures learning potential for early reading development by documenting the amount of assistance needed to learn how to read words with unfamiliar orthography. We examined the additive value of DA for predicting first-grade decoding and word recognition development while controlling for autoregressive effects. Additionally, we examined whether predictive validity of DA would be higher for students who have poor phonological awareness skills. First-grade students ($n = 105$) were assessed on measures of word reading, phonological awareness, rapid automatized naming, and DA in the fall and again assessed on word reading measures in the spring. A series of planned, moderated multiple regression analyses indicated that DA made a significant and unique contribution in predicting word recognition development above and beyond the autoregressor, particularly for students with poor phonological awareness skills. For these students, DA explained 3.5% of the unique variance in end-of-first-grade word recognition that was not attributable to autoregressive effect. Results suggest that DA provides an important source of individual differences in the development of word recognition skills that cannot be fully captured by merely assessing the present level of reading skills through traditional static assessment, particularly for students at risk for developing reading disabilities.

Keywords

dynamic assessment, decoding, word recognition, autoregressor, word reading development

One of the unique purposes of educational assessment is to forecast academic achievement and identify students who may need support, so that instruction can be modified to allow them to achieve their full potential. Conventional ways of forecasting achievement often use a test of already acquired skills for which one wishes to predict (e.g., present level of reading to predict later reading) or a test of precursor skills (e.g., phonological processing for reading). With these assessment methods, students perform without assistance. Thus, these assessments are appropriate for measuring the product of past learning and phenomena that are static in nature. Critics of traditional assessments assert that such static tests do not provide the information needed to meet the goals of the educational assessment. These critics believe that traditional assessments fail to provide information about intra-individual change, are unable to discriminate poor performances due to intrinsic cognitive deficits from a lack of educational opportunities, and do not sample enough items of basic skills, precluding sensitivity for identifying low-performing students (e.g., Daniel, 1997; Sternberg, 1996; Tzuriel & Haywood, 1992).

In response to these criticisms, dynamic assessment (DA) has gained attention as an alternative (for reviews, see Elliott,

2003; Grigorenko & Sternberg, 1998; Jitendra & Kameenui, 1993; Swanson & Lussier, 2001). DA refers to a collection of assessment procedures that allows examiners to intervene with students during a test, either in a form of instruction, via feedback, or in a sequence of progressively explicit prompts. These forms of intervention enable students to make progress when solving difficult problems or mastering novel tasks. Students' learning—that is, responsiveness to DA feedback and instruction—then serves as an indicator of their learning potential. Thus, instruction provided as part of DA allows examiners to capture intra-individual change that is hypothesized to be partially independent from educational history.

¹The Meadows Center for Preventing Educational Risk, University of Texas–Austin, USA

²Florida Center for Reading Research, Florida State University, Tallahassee, USA

³Vanderbilt University, Nashville, TN, USA

⁴Texas State University, San Marcos, USA

Corresponding Author:

Eunsoo Cho, University of Texas Austin, 1912, Speedway, D4900, Austin, TX 78712, USA.

Email: eunsoocho@austin.utexas.edu

The theoretical root of DA can be traced back to the work of Lev Vygotsky (1962) and his notion of the zone of proximal development (ZPD). According to Vygotsky, children's cognitive abilities can be fully understood by recognizing both the actualized and the actualizing abilities. Actualized abilities are those that are completely developed, reflecting what students have learned, whereas actualizing abilities are those that are not yet fully developed but can become actualized in the course of interaction with more advanced individuals. The ZPD is the gap between these two levels. Thus, what DA measures is theoretically different from what static assessments can capture. DA estimates the upper boundary of ZPD, which is how well an individual can learn given assistance (learning potential), whereas static assessments measure the lower boundary of ZPD, which is what has been already learned (learned product). For this reason, DA can offer information about students' future achievement in addition to that provided by static assessments alone.

In this study, we examined the unique contribution of DA in predicting students' development of decoding and word recognition skills during first grade. In addition, we assessed whether DA has higher predictive validity for students at risk for reading disabilities. In what follows, we present an overview of DA literature, review prior work on DA as a predictor of word reading development, and explain how this study extends the literature.

Overview of DA Literature

DA encompasses several different approaches, including learning potential assessment (e.g., Budoff, 1967), testing the limits (Carlson & Wiedl, 1979), mediated assessment (e.g., Feuerstein, Rand, & Hoffman, 1981), learning tests (e.g., Guthke, 1992), and assisted learning and transfer (e.g., Campione, Brown, Ferrara, Jones, & Steinberg, 1985). These approaches differ across various dimensions, including the nature of the interaction (e.g., standardized vs. individualized), the format (e.g., test-teach-retest vs. graduated prompts), and the way learning potential is indexed (e.g., amount of change from pretest to posttest vs. the number of prompts needed to master the skill).

The test-teach-retest format incorporates a blocked scheduling of instruction between pretest and posttest to index the improvement on posttest. Widely used in clinical settings for remediation purposes, individualized instruction tailored to students' needs is embedded. Alternatively, the graduated prompts format uses progressive scheduling of a predetermined hierarchy of prompts and assesses the amount of help students require to master the skill. The level of instructional explicitness is systematically intensified in the graduated prompts format. Because the graduated prompts format has proven to possess sound psychometric properties, it has been widely accepted by

researchers interested in academic achievement for screening and identification of students with special needs (Grigorenko & Sternberg, 1998).

The graduated prompts format was directly influenced by Vygotsky's ZPD concept and information processing theories. Transfer, an ability to apply what has been learned in one context to other contexts, is conceived as synonymous to learning potential (Grigorenko & Sternberg, 1998). In other words, students who can master a set of novel skills with only implicit prompts or incomplete support have higher ability to learn and transfer than students who require explicit instruction. Transfer ability represents an important source of individual differences in academic learning, as students have to learn from instruction and apply that knowledge or skills to different contexts where they have to perform more independently. For example, when students learn how to decode from phonics instruction, they have to apply their decoding skills to reading real words (see Share, 1999). Some students need explicit teaching or extensive practice for their decoding skills to transfer to word recognition skills; others may need only a subtle prompt.

Recently, the value of DA has been underscored as an alternative or supplemental tool to traditional static assessment for predicting students' academic skill development, particularly within response to intervention (RTI) models. In RTI models, students receive increasingly intensive tiers of instruction, depending on their responsiveness to the less intensive instructional tier. If a student fails to respond to the most intensive instruction that the majority of students benefit from, he or she is identified as having a learning disability. In this way, the purpose and procedure of DA and RTI are similar: They take into account students' responsiveness to instruction as an important indicator of individual differences in learning and merge instruction and assessment (Grigorenko, 2009; Wagner & Compton, 2011). Instruction serves as a test in RTI models, whereas a testing procedure is instruction in DA.

Within this context, recent studies have focused on developing DA in the core curriculum (e.g., reading, math). Advocates of curriculum-based DA insist that if an assessment is to be used to identify students with learning disabilities and inform instruction, the test content should be directly linked and applicable to classroom curriculum. A further advantage of aligning DA tasks with the curriculum is that aligned DA measures have the potential to be of higher predictive value than DA measures that use domain-general tasks (e.g., Haywood & Lidz, 2007).

Prior DA Studies on Predicting Word Reading

We are primarily concerned with DA's predictive validity, as we propose DA as a supplementary tool that can aid the process of early identification of students with learning

disabilities in RTI models. To contextualize the present study, we reviewed studies that used DA for predicting word reading development.

Can DA add more? DA has been used to predict later word reading skills, including rate of growth or reading disability status. DA has been developed in a variety of skills within reading, such as phonological awareness (PA; Bridges & Catts, 2011; O'Connor & Jenkins, 1999; Spector, 1992), decoding (Cho, Compton, Fuchs, Fuchs, & Bouton, 2014; Compton et al., 2010; Fuchs, Compton, Fuchs, Bouton, & Caffrey, 2011; Fuchs et al., 2007; Petersen, Allen, & Spencer, 2014), and domains outside of reading such as working memory (Swanson, 2010). All of these studies used DA to measure learning potential, which is an ability to learn from instruction. Although specific DA procedures were slightly different across these studies, the majority included instructional prompts in tiers between multiple learning trials.

Results from prior studies support the incremental value of DA in explaining future word reading outcomes. Swanson (2010) used working memory DA to predict decoding fluency growth across 3 years in middle elementary grades with a sample largely consisting of students at risk for reading and/or math disabilities. Four scores were derived from the working memory DA: Initial (pretest score), Gain (maximum score during teaching), Maintain (posttest score), and Probe (number of prompts during teaching). Gain and Probe scores were significant predictors of decoding skill growth, controlling for disability status. However, because the four scores were entered as a block, it is unknown how much added value Gain and Probe scores (the way we define DA) have independent from static performance. Another limitation of this study is that DA was not compared to other well-known measures of word reading development, such as phonological processing skills.

In other studies, PA tasks were used in DA with kindergarten students. Embedding graduated instructional prompts in phoneme segmentation assessment, Spector (1992) found that the number of prompts uniquely explained 21% of the variance at the end of kindergarten in word recognition, controlling for oral vocabulary and three other static measures of PA. Bridges and Catts (2011) used a DA design similar to Spector's but with phoneme deletion tasks. They used two samples: one a homogeneously Caucasian group with a small percentage of students of low socioeconomic status and the other a diverse group, half of whom were at-risk students. Across these two samples, the researchers found that DA uniquely explained approximately 10% of the variance in later decoding and 5% of the variance in later word recognition, controlling for one other static PA measure. However, neither of the studies tested DA's additive value compared to other possible covariates, possibly overstating the incremental value of DA.

Research that used decoding tasks in DA ruled out the possibility of superfluous predictive value of DA by including a wide range of reading-related and domain-general covariates (Cho et al., 2014; Fuchs et al., 2011). In the DA used in these studies, students were asked to sequentially master three decoding patterns (consonant-vowel-consonant [CVC], CVCe, and CVCCing), using nonwords. Increasingly explicit levels of instruction were provided to help students master each pattern, using onset-rime strategies. The number of instructional levels required for students to master reading each decoding pattern indicated students' learning potential for early reading. In Fuchs et al.'s (2011) study, in which more than half of the sample were at-risk first graders, DA administered at the beginning of first grade predicted timed and untimed word recognition at the end of first grade. When DA was compared against only a static decoding measure, it uniquely predicted 3% to 6% of the variance in the outcome. However, DA predicted only 2.3% of the unique variance in untimed word reading in a model containing measures of language, IQ, attention, PA, and rapid automatized naming (RAN). Cho et al. (2014) extended the study by situating DA in an RTI context and used DA to predict response to Tier 2 reading instruction. They found that DA was predictive of growth during Tier 2 and final word reading intercept at the end of Tier 2, when compared with a series of competing predictors that included Tier 1 slope and intercept, timed and untimed decoding, as well as IQ and phonological processing skills. DA uniquely and significantly explained 3% to 13% of the variance in Tier 2 responsiveness.

It seems clear that the amount of support required to successfully perform reading-related tasks that were initially beyond one's actualized level is associated with unique variance in word reading development beyond the variance accounted for by other static predictors. What has not been explored is whether DA has additive value over and beyond the autoregressor (word reading at an earlier point in time) that represents the lower boundary of ZPD when predicting future word reading outcomes.

For whom can DA add more? To use DA as an identification tool, it must have strong predictive validity for at-risk students, including students with disabilities. So our next question was whether DA has higher predictive validity for at-risk students than for students not at risk.

There are several reasons why DA, compared to static assessments, has the potential to better distinguish at-risk students, especially in reading. First, performance on static assessments of at-risk students can be affected by various factors, such as lack of solid reading instruction, different cultural or linguistic backgrounds, or intrinsic cognitive deficits. Instruction in DA allows examiners to rule out potential confounds surrounding poor performance. In fact, researchers have found that when used as a supplementary

or alternative classification tool, DA improves classification accuracy for students in kindergarten through first grade by reducing the false positives (Compton et al., 2010; O'Connor & Jenkins, 1999; Petersen et al., 2014).

Second, at-risk students are susceptible to floor effects on traditional reading assessments, as these tests do not sensitively capture the growth of students in the lower end of the distribution (see Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009). Traditional standardized reading tests often do not sample enough number of easy items. If a student hits the ceiling quickly on the easiest items on a test, there is no way to seek additional information. On the other hand, learning opportunities provided in DA can offer information about growth. DA allows examiners to seek information from students' failure and gain data about how much scaffolding is needed to help the student succeed. Thus, compared to static assessments, DA has the potential to be more informative for students who are at risk for failure than for students who are already successful.

In a similar vein, static assessments may suffice for average- or high-performing students and DA may not add to the information gathered from the static tests. These students' actualized and actualizing levels may not differ greatly. For example, typically developing students may already perform at their fullest potential on a static assessment because they have had the opportunities to learn and respond appropriately to learning opportunities. At-risk students may not have had these opportunities.

The idea of DA's differential predictive validity has not been empirically tested. In fact, Caffrey, Fuchs, and Fuchs (2008) raised similar questions. In a review of DA studies, the authors categorized the studies into four groups, depending on the target sample: mixed ability, normally achieving, at-risk or disadvantaged, and students with disabilities. Surprisingly, effect sizes for the Pearson correlation between DA and achievement were slightly lower for the at-risk or disadvantaged group, for whom DA is often designed. One limitation of this result is that the correlation effect sizes are reported descriptively, so we cannot statistically compare across groups. In addition, the reported Pearson correlation cannot inform us about the unique additive value of DA to static assessments. To summarize, what we know is that DA has predictive validity for explaining reading development in general. However, for whom DA has stronger additional predictive validity has only been indirectly addressed in prior research.

Present Study

The purpose of the present study was twofold. First, we examined the relative importance of a decoding DA as a predictor of two word reading skills that are important during students' early years of learning to read: decoding (pseudo-word reading) and word recognition (real-word reading). We

extended prior studies by including the autoregressors representing the actualized level of word reading skills at the beginning of first grade for predicting word reading skills at the end of first grade (actualizing level of word reading skills). Our hypothesis was that if DA is a valid measure of learning potential, DA measured in the fall of first grade should explain word reading skills in the spring of first grade over and above what can be explained by word reading skills measured in the fall. Moreover, we added PA and RAN to see whether this relation would hold even after controlling for the important precursors of reading development. Second, we examined whether DA's additional predictive validity differs for students at risk for developing reading disabilities. We considered PA skills, the most common screener for reading disabilities, as a moderator. We hypothesized that stronger predictive validity of DA would exist for students whose PA skills were poor.

Method

Participants

First-grade students with a wide range of reading skills and whose first language is English participated in this study. We initially tested 112 students in the fall (Time 1) and followed up with 105 of the same students in May of first grade (Time 2). The only reason for attrition was moving out of the school district. No differential attrition was found in the demographic information and in the reading measures at Time 1. Demographic information and summary statistics for the final sample are provided in Table 1.

Measures

DA of word reading. The DA of word reading used in this study was modified from the measure developed by Fuchs et al. (2007), which has been shown to have incremental validity for predicting later word reading outcomes (e.g., Cho et al., 2014; Fuchs et al., 2011) and to improve classification accuracy (Compton et al., 2010) when added to a comprehensive set of reading predictors (for a detailed description of the previous DA, see Fuchs et al., 2007). We also incorporated a paired associative learning task into the DA of word reading influenced by Elbro, Daugaard, and Gellert's (2012) DA design.

We developed a DA of word reading that mirrors cognitive skills required for acquiring word reading skills. Three essential skills assessed in the DA are (a) learning symbol-sound correspondence, (b) blending sounds, and (c) inferring decoding rule. We used the following tasks to assess the learning and mastery of each skill. For the learning symbol-sound correspondence task (DA 1), students were asked to make connections between six new symbols (adopted from Chinese characters) and their corresponding

Table 1. Demographics of the Participants ($N = 105$).

Variables	Total Sample Frequencies	Not-at-Risk Frequencies	At-Risk Frequencies
Gender			
Male	58 (55%)	30	28
Female	47 (45%)	24	23
Race			
African American	39 (37%)	11	28
Caucasian	52 (50%)	37	15
Others	13 (12%)	6	7
Missing	1 (1%)	—	1
Free/reduced lunch			
No	50 (48%)	23	17
Yes	54 (51%)	21	33
Missing	1 (1%)	—	1
IEP			
No	94 (90%)	54	40
Yes	11 (10%)	0	11 ^a
Age at Time 1	$M = 6.72, SD = .34$	$M = 6.6, SD = .28$	$M = 6.89, SD = .33$

Note. IEP = individualized education program. At-risk designation was determined post hoc based on the moderated multiple-regression results. At-risk students were $-.25$ SD below the mean, using sample-based z scores of the *Comprehensive Test of Phonological Processing-Elision*.

^aAmong students with IEP in the at-risk group, 5 students were identified as ADHD.

English sounds (/s/, /m/, /t/, /p/, /f/, /a/). For the blending task (DA 2), students were asked to read CVC words written with the new symbols they learned in DA 1. For the inferring decoding rule task (DA 3), students were required to discover the “silent *e*” rule and read CVC*e* words.

We adopted a graduated prompts method, in which a predetermined sequence of brief instructions was embedded within a set of mastery tests. Thus, students were provided with multiple learning trials to master the skill. Each learning trial was composed of two parts: instructional prompts and a six-item posttest. The general procedure was as follows. The tester began with a simple presentation of the required skill. If students failed to master the skill at the first learning trial, the next trial was given with instructional prompts to help students master the skill. If students failed again, they were given the next learning trial with more explicit prompts. Increasingly explicit prompts were given until students reached mastery or until all predetermined prompts were provided. If mastery was achieved, students moved to the next task and received a perfect score for the remaining items that were not administered. If students did not show evidence of learning, even after all the levels were provided, the tester stopped administration. In this case, students received a 0 score for the remaining unadministered tasks. Students were provided with a maximum of nine learning trials for DA 1, four trials for DA 2, and five trials for DA 3.

Descriptions of the increasingly explicit prompts for each DA task are as follows. In DA 1, students were presented with the novel symbols and asked to say the sounds. For the first five trials, students were provided with only corrective feedback (paired associate learning). Then, students were

provided with a keyword representing its sound (e.g., /a/ as in *apple*). Next, students were provided with partial picture clues (e.g., apple-like picture that resembles the /a/ symbol). Next, students were provided with complete picture clues (e.g., apple picture). Finally, students were asked to trace over the symbol with their finger. These prompts were based on findings that paired associate learning is a predictor of word reading development independent from phonemic awareness (Hulme, Goetz, Gooch, Adams, & Snowling, 2007; Litt, de Jong, van Bergen, & Nation, 2013; Warmington & Hulme, 2012) and that constructing embedded picture mnemonics is an effective method to teach students letter-sound correspondence (Ehri, 2014; Ehri, Deffner, & Wilce, 1984).

In DA 2, students were presented with CVC words (e.g., *Sam, fat*) written with the novel symbols. First, the tester read the word; second, the tester modeled blending the three sounds in the word; third, the tester modeled tapping out the sounds by breaking down each sound and then demonstrated blending sounds; and finally, along with the tapping out and blending, students were provided with picture clues to help them remember the sound as they blended the sounds.

In DA 3, students were presented with CVC and CVC*e* words (e.g., *Sam/same, fat/fate*). To represent CVC*e* words, a novel symbol was added to represent *e*. First, the tester read the word; second, the tester tapped out the sounds in the word; third, the tester pointed out the middle sound and instructed students that the middle sound of the CVC*e* word was different from that in the CVC word; fourth, the tester explicitly taught the “silent *e*” rule that when another symbol appears at the end, it changes the sound of the middle letter; finally, the tester provided picture clues to remember

the silent *e* rule. Instructional prompts of DA 2 and DA 3 were based on synthetic phonics.

After each of the instructional prompts, students' learning was assessed by using a posttest. The test comprised six items, and mastery was considered five correct responses. The items repeated across the tests but were presented in random order in each mastery test. Mastery test items were not used for instructional prompts. The outcome measure was the sum of the instructional levels for each task. We also differentiated students who achieved mastery after the last prompt from those who still could not master the skill after receiving all the prompts. For example, if a student scored lower than five out of six on the final mastery test, the DA score (number of prompts) was one more than the actual number of instructional levels received. Internal consistency was .71 across all mastery tests. The administration scripts are provided in the Appendix.

PA. We used the *Comprehensive Test of Phonological Processing: Elision* (Wagner, Torgesen, & Rashotte, 1999) to assess students' PA. Children were asked to say a word and then to say the word after deleting a specified part of the word. According to the manual, the test-retest reliability for ages 5 to 7 exceeds .85

RAN. We used the *Comprehensive Test of Phonological Processing: Rapid Digit Naming* (Wagner et al., 1999) to assess the speed at which students can name two sets of digits displayed in an array of 36 digits in each set. Test-retest reliability exceeded .85 for the first-grade students. The original raw score is the number of seconds that students took to complete two sets, but we divided the raw score by 72 to determine how fast a student can name a single digit. We performed this transformation because it has been shown to have better distributional properties than the original metric (e.g., de Jong, 2011).

Decoding. We used the *Woodcock Reading Mastery Tests-Revised/Normative Update: Word Attack* (Woodcock, 1998) to measure untimed pseudoword reading in isolation. The manual reports the split-half reliability for first-grade students as .94. In addition, we used the *Test of Word Reading Efficiency: Phonemic Decoding Efficiency* (Torgesen, Wagner, & Rashotte, 2012) to measure decoding accuracy and fluency. Test-retest reliability reported in the manual is .86 for the first-grade sample.

Word recognition. We used the *Woodcock Reading Mastery Tests-Revised/Normative Update: Word Identification* (Woodcock, 1998) to measure untimed real-word reading in isolation. The split-half reliability from the manual is .98 for first-grade students. Also, we used the *Test of Word Reading Efficiency: Sight Word Efficiency* (Torgesen et al., 2012) to measure students' word recognition fluency. Test-retest reliability reported in the manual is .93 for the first-grade sample.

Procedures

In October and November of first grade, all of the measures were administered to students during two 1-hour testing sessions and teachers completed a demographic form (including date of birth, ethnicity, free/reduced lunch status as a proxy for socioeconomic status, placement information, English language learner status, and native language). In May of first grade, the decoding and word recognition measures were given again. The testers were trained to follow all administration procedures for the tests. Seven testers were trained to criterion, using standard directions for administration. All individual sessions were audiotaped; we randomly selected 20% of the tapes, stratifying by tester, to check for procedural fidelity and scoring reliability by an independent scorer. Procedural fidelity was above 97% across all assessments, and scoring reliability exceeded 90% agreement.

Data Analyses and Results

Descriptive Characteristics of the Sample and Data Issues

Descriptive statistics and zero-order correlation are presented in Table 2. To provide a basis for comparison to the norm, we present raw scores as well as standard scores when applicable. Although means of Word Attack and Word Identification standard scores are high, these scores are inflated by using the norms of the *Woodcock Reading Mastery Tests-Revised/Normative Update*, published in 1998. This inflation is not unique in our sample, as it was reported in Bridges and Catt's (2011) study. Based on the *Test of Word Reading Efficiency* norms obtained in 2008–2009, we consider the students' reading skills to be comparable to their normative sample—thus, representative of the population.

We initially standardized the variables, using our sample's statistics, because there were large scaling differences across measures, as can be seen in Table 2. Once all the measures were *z*-scored, we derived composite scores for decoding and word recognition at both time points by averaging the *z*-scores from Word Attack and Phonemic Decoding Efficiency for decoding and Word Identification and Sight Word Efficiency for word recognition. In addition, we created a DA composite score by averaging the *z*-scores from DA 1 to DA 3.

Prior to analysis, data (raw scores) were screened for outliers and univariate and multivariate normalities. No extreme values were detected, using the $M \pm 2.5 SD$ criterion. But we found the following variables to violate the assumptions of univariate normality: Both word recognition measures at Time 2 had a significant negative skew (Sight Word Efficiency = $-.88$; Word Identification = $-.97$), and PA showed a significant positive skew (.68). However, when criteria of values greater than $|2|$ for skewness and $|7|$ for kurtosis were used, as suggested by West, Finch, and

Table 2. Descriptive Statistics and Zero-Order Correlation.

Measures	Time 1									Time 2			
	DA 1	DA 2	DA 3	PA	RAN	WID	SWE	WAT	PDE	WID	SWE	WAT	PDE
Time 1													
DA 1	—												
DA 2	.40	—											
DA 3	.23	.39	—										
PA	-.14	-.34	-.39	—									
RAN	-.11	-.25	-.20	.02	—								
WID	-.21	-.24	-.43	.55	.33	—							
SWE	-.10	-.24	-.27	.42	.44	.89	—						
WAT	-.22	-.43	-.55	.64	.25	.86	.73	—					
PDE	-.24	-.31	-.38	.50	.39	.83	.83	.81	—				
Time 2													
WID	-.22	-.45	-.46	.60	.38	.91	.81	.82	.80	—			
SWE	-.11	-.34	-.32	.42	.50	.78	.84	.64	.70	.86	—		
WAT	-.22	-.42	-.47	.50	.39	.78	.69	.82	.77	.84	.69	—	
PDE	-.19	-.39	-.43	.50	.45	.80	.81	.78	.85	.86	.81	.85	—
Raw score													
M	4.98	1.98	3.7	8.53	1.41	39.49	34.55	16.30	14.06	50.39	45.42	21.10	19.08
SD	2.86	1.52	2.11	4.52	.33	13.87	14.11	8.16	8.49	14.02	14.60	9.25	10.10
Standard score													
M	—	—	—	10.98	—	118.70	105.79	117.36	100.09	114.01	106.41	113.09	101.28
SD	—	—	—	3.01	—	10.67	16.16	12.27	15.61	11.00	18.20	12.20	15.49
Range													
Min.	1	1	1	0	.74	2	5	0	0	2	5	0	0
Max.	10	5	6	19	2.4	71	67	35	45	79	72	39	41

Note. DA = dynamic assessment; PA = phonemic awareness; PDE = *Test of Word Reading Efficiency: Phonemic Decoding Efficiency*; RAN = rapid automatized naming; SWE = *Test of Word Reading Efficiency: Sight Word Efficiency*; WAT = *Woodcock Reading Mastery Tests–Revised/Normative Update: Word Attack*; WID = *Woodcock Reading Mastery Tests–Revised/Normative Update: Word Identification*. Correlation coefficients greater than .19 are significant at an alpha level of .05. RAN does not have a standard score because digit naming is not normed for 6-year-old students.

Curran (1995), there was no indication of large floor or ceiling effects for any of the measures. We also detected multivariate skewness based on the Mardia’s statistics (5.88), $\chi^2(35) = 106.67$. In addition, we found relations between PA and other word reading variables to deviate from linearity. Preliminary regression analyses also resulted in four observations with studentized residuals greater than 3, nonnormal error distribution, and heteroscedasticity.

To address these data issues, we first used standard scores for PA because they had better distributional characteristics and resolved the issue of nonlinearity. Raw score distribution showed that the majority of students scored between 8 and 9. Standard score conversion helped with dispersing those students and making distribution look more normal. Second, we used different robust regression estimators to see the effects of these violations. We ran robust regression models by using MM-estimator, a robust estimator developed by Yohai (1987), to take into account issues concerning effects of outliers, heterogeneity, and lack of normality. MM-estimator is the most commonly employed robust regression technique (Andersen, 2008). It has a high breakdown point in that the estimates are stable

up to the point where 50% of the data are contaminated (e.g., outliers). It also uses iteratively reweighting least squares, giving less weight to outliers until convergence. The results are then compared with a method that corrects for minor nonnormality and heteroscedasticity by using robust standard errors (also known as the Huber-White sandwich estimator). We found identical patterns in the directions and the significance of the coefficients between the two methods—thus, we provide results from the latter that are more easily interpreted and widely used.

Moderated multiple regression. We ran a series of planned multiple-regression analyses with the following steps for the base models and the extended models. For the base models, we compared DA only to the autoregressor by initially running a simple regression model with the autoregressor and then adding DA to the model. For extended models, we initially entered the autoregressor, PA, and RAN in the first step; then added DA in the second step; and finally included the $PA \times DA$ interaction term in the third step to see whether DA has differential predictive value, depending on the student’s initial PA level. The interaction term was entered into

Table 3. Multiple-Regression Results.

Models/Steps	Word Recognition						Decoding					
	Beta	SE	t	p	R ²	ΔR^2	Beta	SE	t	p	R ²	ΔR^2
Base models												
Step 1					.80						.78	
Autoregressor	.89	.06	15.93	.00			.89	.05	16.88	.00		
Step 2					.82	.02*					.78	.00
Autoregressor	.85	.05	45.69	.00			.85	.06	14.38	.00		
DA	-.13	.06	-3.11	.00			-.06	.06	-1.34	.18		
Extended models												
Step 1					.83						.80	
Autoregressor	.75	.05	15.10	.00			.79	.07	11.92	.00		
PA	.16	.04	3.64	.00			.05	.06	.81	.42		
RAN	.15	.04	3.32	.00			.17	.04	3.72	.00		
Step 2					.84	.01*					.80	.00
Autoregressor	.74	.05	15.07	.00			.78	.07	11.30	.00		
PA	.13	.04	2.98	.00			.05	.07	.69	.50		
RAN	.13	.04	3.05	.00			.17	.04	3.64	.00		
DA	-.09	.06	-2.08	.04			-.04	.09	-.89	.34		
Step 3					.85	.01*					.81	.00
Autoregressor	.75	.05	14.92	.00			.78	.07	11.81	.00		
PA	.11	.04	2.64	.01			.04	.07	.56	.58		
RAN	.11	.04	2.46	.02			.15	.04	3.24	.00		
DA	-.08	.06	-1.77	.08			-.03	.06	-.76	.45		
DA × PA	.12	.04	3.17	.00			.06	.05	1.69	.09		

* $p < .05$.

the regression equation as planned regardless of the results from the previous step because interaction effects can be found in the absence of the average effect. Results are presented in Table 3.

As expected, the autoregressor was the most powerful predictor of word reading development. DA neither was significant in the base model nor did it show differential predictive value in the extended model for explaining decoding outcome. For word recognition, even in the presence of the autoregressor, DA was a significant predictor ($\beta = -.13$, $t = -3.11$, $p < .05$), explaining a small but significant amount of additional variance (2%). For the extended model, the autoregressor, PA, and RAN were all significant predictors of word recognition at Time 2. DA was significant even after controlling for the strongest set of predictors ($\beta = -.09$, $t = -2.08$, $p < .05$) but explained less than 1% of the variance, although significant. Lastly, we added the PA × DA interaction and found a significant moderating effect of PA ($\beta = .12$, $t = 3.17$, $p < .005$). All other predictors were still significant, except for the marginal effect of DA. Adding the interaction term improved the model by explaining an additional 1.28% of the variance in word recognition development. To further understand the pattern of interaction, we probed the marginal effects of DA to see the region of significance along the value of PA. When controlling for the autoregressive effects and RAN, DA was predictive of word recognition at Time 2 for students whose z -score of PA was at or below $-.25$ (see Figure 1).

To supplement these regression results, a set of commonality analyses was conducted to determine the unique and common contributions between DA and the autoregressor for predicting word recognition development. We conducted a separate commonality analysis for the two groups based on the results from the moderated multiple-regression plots (see Table 4). We created at-risk (low-PA) and not-at-risk (high-PA) groups using a cut score of $-.25$ z -scored PA. Low-PA students had a mean of 8.7 in scale score of PA, which has a mean of 10 and SD of 3 (see Table 5). We found that the total variance explained by DA was greater for the not-at-risk group (16%) than for the at-risk group (10%). However, when we look at the additive effects of DA to the autoregressor, DA uniquely accounted for 1.70% of the explained variance for the not-at-risk group, whereas a greater amount of variance, 3.49%, was attributable uniquely to DA for the at-risk group.

Discussion

The purpose of this study was to assess the value of DA in predicting decoding and word recognition development, controlling for the present level of word reading skills (i.e., the autoregressor). In this way, we contrasted the ZPD against the zone of actual development for predicting word reading development. We also examined whether DA's additional predictive value to the autoregressor is different across varying levels of PA skills. We specifically hypothesized that

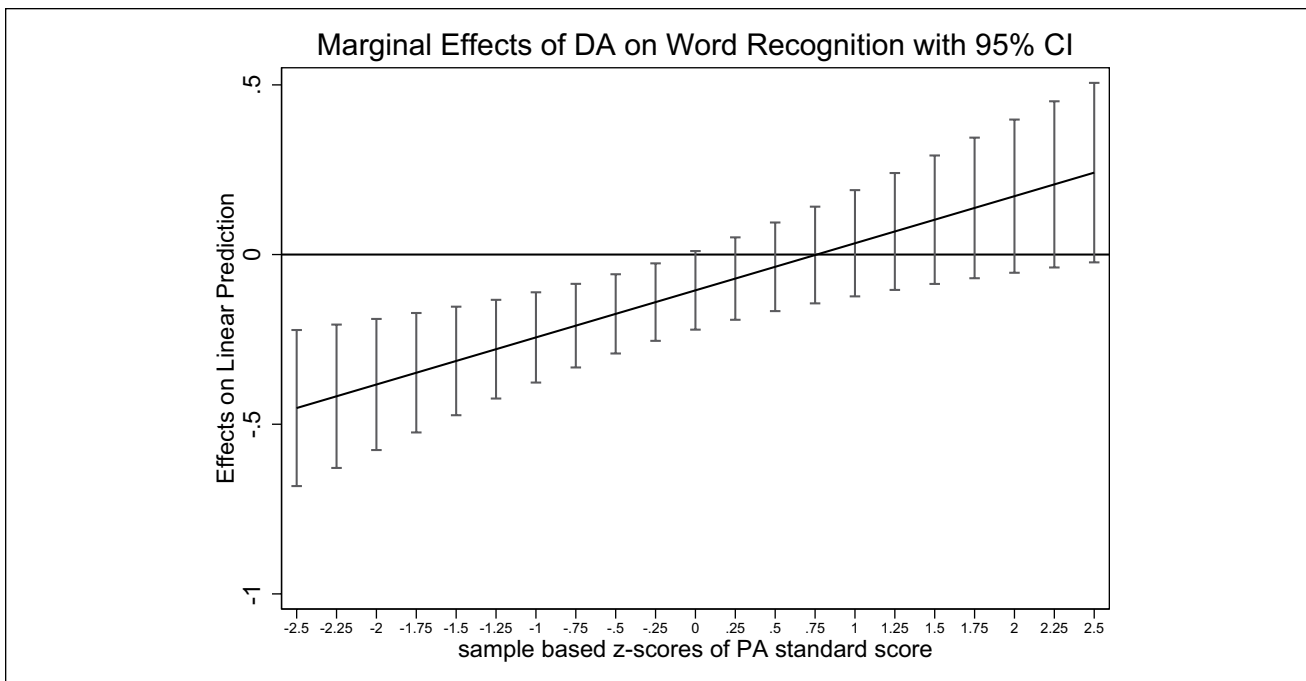


Figure 1. Marginal effects of dynamic assessment on end-of-first-grade word recognition at varying levels of initial phonemic awareness score.

Table 4. Commonality Analysis for Predicting End-of-First-Grade Word Recognition Development by Initial PA Level.

Variables	Word Recognition						Decoding	
	Total		At Risk		Not at Risk		Total	
	Coeff.	%	Coeff.	%	Coeff.	%	Coeff.	%
Unique to autoregressor	.62	76.15	.71	88.01	.62	79.13	.55	69.94
Unique to DA	.02	1.91	.03	3.49	.01	1.70	.00	.39
Common to autoregressor DA	.18	21.94	.07	8.50	.15	19.17	.23	29.67
Total	.82	100	.81	100	.78	100	.78	100

Note. Coeff. = coefficient (proportion of variance explained).

DA would show greater utility for predicting word reading development for at-risk students with poor PA skills.

Can DA Add More?

Prior studies have consistently shown that DA has predictive validity for forecasting word reading development, controlling for the statically measured precursors of reading such as PA and RAN. In this study, we extended the literature by comparing DA to the autoregressor, thereby putting DA to the most stringent test. Longitudinal studies have shown high stability of word reading development from kindergarten to third grade in that year-to-year word reading correlations are above .7 and even up to .9 (Parrila, Kirby, & McQuarrie, 2004; Wagner et al., 1997). When translated to R^2 , the autoregressor explains approximately

50% to 80% of the variance in later word reading. Thus, 20% to 50% of the variance in future word reading comprises random errors and/or systematic variations. We found that DA, on average, predicts first-grade students' word recognition development and explains an additional 2% of the variance in word recognition development. Although 2% seems small, given the fact that our prediction occurred in a shorter period (within first grade) than the other studies, we assume that the autoregressive effects might have been stronger in this study than in prior studies and, hence, the effects of DA should not be trivialized. To supplement the understanding of DA's value compared to PA or RAN, we examined R^2 change from the model with only the autoregressor to the model with the autoregressor and either PA or RAN. In the current sample, R^2 change was .01 for PA and

Table 5. Descriptive Statistics and Zero-Order Correlation for Not-At-Risk (N = 54) and At-Risk (N = 51) Groups.

Measures	Time 1										Time 2										Raw Score				Standard Score	
	DA 1	DA 2	DA 3	PA	RAN	WID	SWE	WAT	PDE	WID	SWE	WAT	PDE	WAT	PDE	M	SD	M	SD	M	SD					
Time 1																										
DA 1	—	.46	.20	-.16	-.03	-.24	-.15	-.18	-.18	-.28	-.17	-.20	-.13	—	—	4.54	2.77	—	—	—	—					
DA 2	.32	—	.38	-.12	-.24	-.38	-.36	-.44	-.26	-.48	-.42	-.36	-.29	—	—	1.53	1.11	—	—	—	—					
DA 3	.27	.29	—	-.19	-.04	-.29	-.10	-.39	-.14	-.21	-.13	-.20	-.18	—	—	3.00	2.08	—	—	—	—					
PA	-.27	-.34	-.29	—	-.18	.23	.12	.31	.20	.32	.07	.13	.20	—	—	11.39	4.21	13.11	2.29	—	—					
RAN	-.18	-.27	-.39	.35	—	.26	.43	.19	.33	.24	.47	.33	.43	—	—	1.41	.32	—	—	—	—					
WID	-.08	-.15	-.34	.62	.47	—	.82	.80	.78	.89	.65	.60	.65	—	—	46.02	11.24	123.52	8.60	—	—					
SWE	.05	-.01	-.22	.55	.51	.91	—	.65	.84	.76	.81	.57	.79	—	—	39.52	12.75	111.63	15.11	—	—					
WAT	-.14	-.26	-.52	.62	.41	.84	.71	—	.66	.74	.53	.66	.60	—	—	21.04	6.16	123.78	7.05	—	—					
PDE	-.19	-.19	-.44	.59	.54	.81	.77	.88	—	.77	.67	.62	.80	—	—	17.56	8.06	103.61	14.83	—	—					
Time 2																										
WID	-.09	-.32	-.41	.70	.55	.90	.81	.78	.77	—	.73	.72	.76	—	—	56.81	9.84	118.91	7.32	—	—					
SWE	.00	-.23	-.32	.66	.58	.83	.84	.63	.68	.91	—	.51	.79	—	—	50.02	11.30	112.48	19.09	—	—					
WAT	-.13	-.31	-.53	.57	.53	.80	.69	.85	.83	.85	.72	—	.71	—	—	25.90	7.31	118.43	8.13	—	—					
PDE	-.15	-.30	-.44	.52	.57	.82	.77	.79	.85	.86	.80	.88	—	—	—	23.56	8.28	109.48	11.15	—	—					
Raw score																										
M	5.45	2.45	4.51	5.51	1.41	32.57	29.29	11.27	10.35	43.58	40.55	16.67	14.33	—	—	—	—	—	—	—	—					
SD	2.89	1.75	1.86	2.37	.35	13.09	13.67	6.95	7.34	14.64	16.15	9.08	9.74	—	—	—	—	—	—	—	—					
Standard score																										
M	—	—	—	8.73	—	113.59	99.60	110.57	96.35	108.82	100.00	107.43	92.58	—	—	—	—	—	—	—	—					
SD	—	—	—	1.8	—	10.32	15.70	13.00	15.69	11.90	14.87	13.27	14.76	—	—	—	—	—	—	—	—					

Note. Correlation above diagonal is for not-at-risk students, and correlation below diagonal is for at-risk students. Correlation coefficients greater than .23 for the not-at-risk group and at .24 for the at-risk group are significant at an alpha level of .05.

.02 for RAN. Thus, DA's additive value to the autoregressor was comparable to that of phonological processing abilities.

The average additive effect of DA on word recognition development was significant in the extended model with PA and RAN. This finding suggests that DA's predictive validity was not due to phonological processing abilities involved in performing DA tasks. As evidenced by the significant correlations DA has with PA, RAN, and the autoregressor, phonological skills play an important role in all DA tasks (acquiring symbol-sound correspondence, blending sounds, and figuring out the rule of association between sounds and orthography). However, it was not only the phonological aspect of DA that made it predictive of word recognition development. Presumably, it was the "learning" aspect of DA that was predictive of variance unexplained by the static assessments. Results suggest that although learning potential of early reading is not independent of phonological processing skills, good phonological processing skills are not sufficient for comprehensively understanding students' early reading learning potential. Our conclusion is that DA of word reading captures the actualizing word reading skills (learning potential) at the beginning of first grade and allows us to forecast students' word reading skills that will eventually be actualized.

For Whom Can DA Add More?

The role of DA becomes especially meaningful for students at risk for developing reading disabilities. When we allowed predictive validity of DA to differ across students' PA level, DA was a significant predictor of word recognition development only for students with poor PA skills. Examining the marginal effects of DA, we found that DA had predictive validity for students whose PA score was more than .25 *SD* below the sample-specific mean. When converted to raw scores, the majority of these students (94%) scored below 8, which we found to be qualitatively meaningful. On the *Comprehensive Test of Phonological Processing*: Elision task, items are ordered by difficulty. Students are asked to delete a part of compound words in Items 1 through 3, to delete the initial phoneme in one-syllable words in Items 4 through 7, and to delete middle phonemes from Item 8. Thus, if we assume that students who scored 7 answered the first seven items correctly, DA is predictive of word recognition development for students who have not yet mastered deleting middle phonemes. This result coincides with prior DA studies with PA that showed predictive validity on word reading development for kindergarten students for whom static PA assessments typically exhibited floor effects (Bridges & Catts, 2011). What we learned from this study is that DA can be a useful supplemental tool for understanding word recognition development of at-risk students when statically measured PA

does not provide much information about prospective word reading development.

Another interesting finding is the substantial shared variance between the autoregressor and DA for predicting word recognition development for not-at-risk students versus the much smaller proportion of shared variance for at-risk students. This finding suggests that not-at-risk students may already perform to their fullest potential on static assessments and that DA may not provide further information. However, standardized word reading assessments may not fully capture how well at-risk students who perform poorly on the test will read in the future. This finding is in line with the rationale we provided in the introduction for why we hypothesized differential predictive validity.

Our results have implications for RTI models as well. There has been a concern in the field that a standard RTI model, in which more intensive tiers of instruction are provided when students are unresponsive to the previous tiers, could become another "wait to fail" model (Vaughn, Denton, & Fletcher, 2010). For example, when using RTI for identification purposes, Compton et al. (2012) showed that Tier 2 progress-monitoring data, beyond Tier 1 response data and norm-referenced assessments, were not necessary in accurately classifying nonresponders to Tier 2 instruction. They suggested that some students can be identified early as needing Tier 3 intervention without spending several weeks in Tier 2 intervention. From a prevention perspective, Al Otaiba et al. (2014) compared a standard RTI model to a dynamic RTI procedure where students received Tier 2 or Tier 3 immediately based on screening results and found dynamic RTI to be more effective than the standard RTI process. These researchers suggest that not all students need to go through Tier 1 and Tier 2.

Within this context, we propose DA to be one possible way that might help us accurately identify students for each tier. Our results suggest DA to have additional predictive validity beyond static reading measures for at-risk students who might eventually need Tier 3 instruction. DA has been shown to be predictive of Tier 2 responsiveness and to improve classification accuracy, especially when used in a second stage of screening to reduce false positives (Cho et al., 2014; Compton et al., 2010). Our results support the finding of prior studies in that DA does not need to be administered to all students—only to students that static measures identified in the first stage of screening as being at risk (Fuchs, Fuchs, & Compton, 2012).

Decoding and Word Recognition Contrast

Another interesting result is the contrast between the two word reading outcomes. Our DA simulated the learning process of acquiring decoding skills. However, DA was not predictive of decoding development in the presence of

the autoregressor, although there was a high correlation with decoding at Time 2. In contrast, a prior study that used the same DA for predicting concurrent decoding and word recognition outcomes showed the opposite pattern of results: DA was predictive of concurrent decoding but not word recognition skills, controlling for precursors of reading and domain-general learning measures (Cho & Compton, 2015). Commonality analyses with a full sample showed that the shared variance between DA and static decoding measures accounted for approximately 30% of the explained variance in decoding development.

The reason for this inconsistency may be attributed to the fact that learning and transfer is the key concept of the graduated prompts approach (Campione & Brown, 1990). Decoding skill provides a platform in which accurate and automatic word recognition skill develops. Repeated practice is critical in binding the orthographic and phonological representations of a word. As such, substantial time for practice is needed for decoding skills to transfer to word recognition skills (Share, 1995). This fact could also explain why our DA did not have incremental validity for the concurrent word recognition outcome in the prior study but did for future word recognition in the present study. Interestingly, Seethaler, Fuchs, Fuchs, and Compton (2012) found that DA of balancing equations that used graduated prompts did not have incremental predictive value for future outcomes closely related to the DA tasks (computation) but did have incremental validity for an outcome that was more distal to the DA tasks and required students to transfer (word problem). Thus, DA may work better when the outcome is a transferred skill of the DA tasks for predictive validity.

Limitations of This Study and Future Directions

In closing, we note that the present study is not without limitation. First, our sample size was not large enough to include domain-general learning abilities as competing predictors. In a previous study, DA was found to be predictive of reading in the presence of domain-general learning abilities such as nonverbal reasoning and attention (Cho & Compton, 2015). In fact, we acknowledge that other cognitive abilities such as working memory may also be important to perform DA tasks, especially for blending and rule-based learning, where students have to hold symbol-sound knowledge in their working memory. Thus, future studies conducted with larger sample sizes can put DA to the most stringent test by including these domain-general predictors.

Second, examining DA of decoding's predictive validity in beginning kindergarten would be an interesting step in this line of research. Many studies have examined the predictive validity of DA in phonological skills with kindergarten students, but DA of decoding has not been field tested with this population. Having DA tasks more aligned to actual reading process may provide higher predictive validity than DA in phonological skills for kindergarten or even younger children. Another reason for the need of field testing DA of decoding with younger students is to reduce the unintended transfer effects. Although DA is supposed to measure transfer of learned skills during DA, we noted that DA also captured transfer in a way that we did not intend. For example, students who already learned to read words in English and had good transfer ability quickly found similarities between English and newly learned orthography.

Appendix. Instructional Prompts for DA.

Introduction

Hi, my name is _____. Today, we're going to learn how to read words that people from another planet use. These people use funny letters. Let's learn how to read one of their books. Try to see if you can read their words the way they do. Your work will not be part of your grade, but I want you to work really hard and pay careful attention to what I say.

DA 1: Letter-Sound Correspondence Task

DA 1: Levels 1–5 (Paired Associate Learning)

I will show you funny letters and say the sounds that each funny letter makes. Listen carefully and try to remember what sound each funny letter makes.

This says /m/. What sound?

This says /p/. What sound?

This says /f/. What sound?

This says /s/. What sound?

This says /t/. What sound?

This says /a/. What sound?

山
人
魚
日
占
石

(continued)

Appendix (continued)

Mastery Test, DA 1: Levels 1–5

Tell me the sounds of these letters, starting from top to bottom and from left to right. (Immediate corrective feedback is provided.)

DA 1: Level 6

Now I will tell you the sounds each funny letter makes. And I'll give you keywords for each funny letter.

This says /m/ as in *mountain*. What sound? What word?

This says /p/ as in *person*. What sound? What word?

This says /f/ as in *fish*. What sound? What word?

This says /s/ as in *sun*. What sound? What word?

This says /t/ as in *top*. What sound? What word?

This says /a/ as in *apple*. What sound? What word?

山
人
魚
日
占
石

Mastery Test, DA 1: Level 6

Tell me the sounds of these letters. (No feedback is provided.)

DA 1: Level 7

Now I will give you picture clues to remember.

This says /m/ as in *mountain*. Do you know why? Look at this picture. This is an easy way of drawing a mountain. And it looks like the funny letter that says /m/. So this says /m/ as in *mountain*. What sound? What word? (point to each picture)



This says /p/ as in *person*. Do you know why? Look at this picture. This is an easy way of drawing a person. And it looks like the funny letter that says /p/. So this says /p/ as in *person*. What sound? What word? (point to each picture)



This says /f/ as in *fish*. Do you know why? Look at this picture. This is an easy way of drawing a fish. And it looks like the funny letter that says /f/. So this says /f/ as in *fish*. What sound? What word? (point to each picture)



This says /s/ as in *sun*. Do you know why? Look at this picture. This is an easy way of drawing a sun. And it looks like the funny letter that says /s/. So this says /s/ as in *sun*. What sound? What word? (point to each picture)



This says /t/ as in *top*. Do you know why? Look at this picture. This is an easy way of drawing a top. And it looks like the funny letter that says /t/. So this says /t/ as in *top*. What sound? What word? (point to each picture)



This says /a/ as in *apple*. Do you know why? Look at this picture. This is an easy way of drawing an apple. And it looks like the funny letter that says /a/. So this says /a/ as in *apple*. What sound? What word? (point to each picture)



Mastery Test, DA 1: Level 7

Tell me the sounds of these letters. (No feedback is provided.)

DA 1: Level 8

I will give you more helpful clues to remember these sounds.

This is a mountain. And this is an easy way of drawing a mountain. Now, do you see why this makes the /m/ sound as in *mountain*? Because it came from the shape of a mountain, this says /m/ as in *mountain*. What sound? What word?



This is a person. And this is an easy way of drawing a person. Now, do you see why this makes the /p/ sound as in *person*? Because it came from the shape of a person, this says /p/ as in *person*. What sound? What word?



This is a fish. And this is an easy way of drawing a fish. Now, do you see why this makes the /f/ sound as in *fish*? Because it came from the shape of a fish, this says /f/ as in *fish*. What sound? What word?



This is a sun. And this is an easy way of drawing a sun. Now, do you see why this makes the /s/ sound as in *sun*? Because it came from the shape of a sun, this says /s/ as in *sun*. What sound? What word?



This is a top. And this is an easy way of drawing a top. Now, do you see why this makes the /t/ sound as in *top*? Because it came from the shape of a top, this says /t/ as in *top*. What sound? What word?



Appendix (continued)

This is an apple. And this is an easy way of drawing an apple. Now, do you see why this makes the /a/ sound as in *apple*? Because it came from the shape of an apple, this says /a/ as in *apple*. What sound? What word?



Mastery Test, DA 1: Level 8

Tell me the sounds of these letters. (No feedback is provided.)

DA 1: Level 9

This time, I want you to use your finger to trace over the letter and say the sounds.

This says /m/ as in *mountain*. Now, use your finger to trace over the letter. What sound?



This says /p/ as in *person*. Now, use your finger to trace over the letter. What sound?



This says /f/ as in *fish*. Now, use your finger to trace over the letter. What sound?



This says /s/ as in *sun*. Now, use your finger to trace over the letter. What sound?



This says /t/ as in *top*. Now, use your finger to trace over the letter. What sound?



This says /a/ as in *apple*. Now, use your finger to trace over the letter. What sound?



Mastery Test, DA 1: Level 9

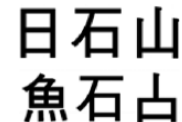
Tell me the sounds of these letters. (No feedback is provided.)

DA 2: Blending Sounds (CVC)

DA 2: Level 1

Because you learned your funny letter sounds, it is time to put together the sounds to make the words you know.

Sam. Your turn. What word?



Fat. Your turn. What word?

Mastery Test, DA 2: Level 1

Read these words to me. (No feedback is provided.)

DA 2: Level 2

Let's try some more. I will show you how to read these words. I will stretch out the sounds in the word and say them fast.

s-a-m, *Sam*. (use your index finger) Your turn.



f-a-t, *fat*. (use your index finger) Your turn.

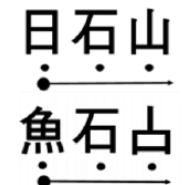
Mastery Test, DA 2: Level 2

Read these words to me. (No feedback is provided.)

DA 2: Level 3

Let's try some more. This time, I am going to tap out the sounds in the word and say them fast.

s.a.m., s-a-m, *Sam*. (use your index finger) Your turn.



f.a.t., f-a-t, *fat*. (use your index finger) Your turn.

Mastery Test, DA 2: Level 3

Read these words to me. (No feedback is provided.)

DA 2: Level 4

Let's try some more. This time, I used the letters with pictures related with the keywords of its sound.

/s/ as in *sun*, /a/ as in *apple*, /m/ as in *mountain*.

s.a.m., s-a-m, *Sam*. (use your index finger) Your turn.



Appendix (continued)

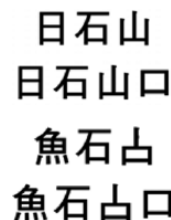
/f/ as in *fish*, /a/ as in *apple*, /t/ as in *top*.
 f.a.t., f-a-t, *fat*. (use your index finger) Your turn.



Mastery Test, DA 2: Level 4
 Read these words to me. (No feedback is provided.)

DA 3: Rule-Based Learning (CVCe)

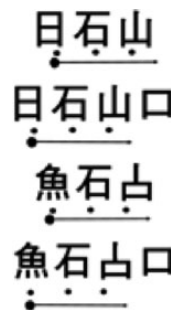
DA 3: Level 1
 Because you are doing a good job working hard, let's try something new.
 Sam. What word?
 Same. What word?
 (point to each word)



Fat. What word?
 Fate. What word?
 (point to each word)

Mastery Test, DA 3: Level 1
 Read these words to me. (No feedback is provided.)

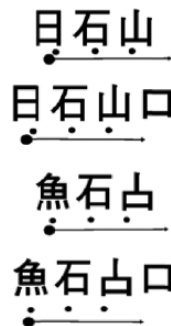
DA 3: Level 2
 This time, I am going to tap out each sound and say them fast.
 s.a.m., s-a-m, *Sam*. Your turn.
 s.ā.m., s-ā-m, *same*. Your turn.
 (point to each word)



f.a.t., f-a-t, *fat*. Your turn.
 f.ā.t., f-ā-t, *fate*. Your turn.
 (point to each word)

Mastery Test, DA 3: Level 2
 Read these words to me. (No feedback is provided.)

DA 3: Level 3
 I will tap out each sound and say them fast again. This time, listen to the middle sound to see how it changes.
 s.a.m., *Sam*. Your turn. s.ā.m., *same*. Your turn. (point to each word) Does this letter in *Sam* and *same* say the same sound?
 (point to middle sound) No, this letter in *Sam* says the /a/ sound as in *apple*. But this letter in *same* says the /ā/ sound as in *apricot*.



f.a.t., *fat*. Your turn. f.ā.t., *fate*. Your turn. (point to each word) Does this letter in *fat* and *fate* say the same sound?
 (point to middle sound) No, this letter in *fat* says the /a/ sound as in *apple*. But this letter in *fate* says the /ā/ sound as in *apricot*.

Mastery Test, DA 3: Level 3
 Read these words to me. (No feedback is provided.)

DA 3: Level 4
 Now, I will tell you why this letter makes different sounds.
 This last funny letter in *same* does not have a sound. Instead, it changes the sound of the middle letter. We call this the magic square because it changes the middle sound.
 s.a.m., *Sam*. Your turn. See this magic square? Listen carefully. s.ā.m., *same*. Your turn.
 (point to each word)



Appendix (continued)

This last funny letter in *fate* does not have a sound. Instead, it changes the sound of the middle letter. We call this the magic square because it changes the middle sound.
f.a.t., *fat*. Your turn. See this magic square? Listen carefully. f.ā.t., *fate*. Your turn.
(point to each word)



Mastery Test, DA 3: Level 4
Read these words to me. (No feedback is provided.)

DA 3: Level 5
Now, I will give you the keyword and picture to help you remember.
This is a picture of an apricot.
This is a funny letter that comes from the shape of an apricot. So it says /ā/ as in *apricot*.
This looks like this funny letter /a/, but it has a big seed inside like an apricot.



When there is a magic square, and it changes the middle sound, /a/ as in *apple* becomes /ā/ as in *apricot*. Like this!
So this says s.ā.m., *same*.
Let's try.
s.a.m., *Sam*. Your turn.
s.ā.m., *same*. Your turn.



When there is a magic square and it changes the middle sound, /a/ as in *apple* becomes /ā/ as in *apricot*. Like this!
So this says f.ā.t., *fate*.
Let's try.
f.a.t., *fat*. Your turn.
f.ā.t., *fate*. Your turn.



Mastery Test, DA 3: Level 5
Read these words to me. (No feedback is provided.)

Thus, to reduce this unintended transfer effect and to minimize the influence of prior reading skills, using DA with younger students who have not learned to read would provide information about how DA functions as a test of learning potential.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100013 to The University of Texas at Austin as part of the Reading for Understanding Research Initiative and by Grants R324G060036 and R305A100034 from the Institute of Education Sciences (IES) in the U.S. Department of Education, and by Core Grant HD15052 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development

(NICHD) to Vanderbilt University. Study data were entered and managed using REDCap electronic data capture tools hosted at Vanderbilt University, which was supported by Vanderbilt Institute for Clinical and Translational Research Grant UL1 TR000445 from NCATS/NIH. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

- Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in Tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children, 81*, 11–27.
- Andersen, R. (2008). *Modern methods for robust regression* (Issue 152). Thousand Oaks, CA: Sage.
- Bridges, M. S., & Catts, H. W. (2011). The use of a dynamic screening of phonological awareness to predict risk for reading disabilities in kindergarten children. *Journal of Learning Disabilities, 44*, 330–338.
- Budoff, M. (1967). Learning potential among institutionalized young adult retardates. *American Journal of Mental Deficiency, 72*, 404–411.

- Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *The Journal of Special Education, 41*, 254–270.
- Campione, J. C., & Brown, A. L. (1990). Guided learning and transfer: Implications for approaches to assessment. In N. Frederiksen & R. Glaser (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 141–172). Hillsdale, NJ: Lawrence Erlbaum.
- Campione, J. C., Brown, A. L., Ferrara, R. A., Jones, R. S., & Steinberg, E. (1985). Breakdowns in the flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence, 9*, 297–315.
- Carlson, J. S., & Wiedl, K. H. (1979). Toward a differential testing approach: Testing-the-limits employing the Raven matrices. *Intelligence, 3*, 323–344.
- Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42*(2), 163–176. doi:10.1177/0022219408326219
- Cho, E., & Compton, D. L. (2015). Construct and incremental validity of DA within and across domains. *Learning and Individual Differences, 37*, 183–196.
- Cho, E., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (2014). Examining the predictive validity of a dynamic assessment of decoding to forecast response to Tier 2 intervention. *Journal of Learning Disabilities, 47*, 409–423.
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., & Crouch, R. C. (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*, 327–340.
- Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., & Bouton, B. (2012). Accelerating chronically unresponsive children to Tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities, 45*, 204–216.
- Daniel, M. H. (1997). Intelligence testing: Status and trends. *American Psychologist, 52*, 1038–1045.
- de Jong, P. F. (2011). What discrete and serial rapid automatized naming can reveal about reading. *Scientific Studies of Reading, 15*, 314–337.
- Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18*, 5–21.
- Ehri, L. C., Deffner, N. D., & Wilce, L. S. (1984). Pictorial mnemonics for phonics. *Journal of Educational Psychology, 76*, 880–893.
- Elbro, C., Daugaard, H. T., & Gellert, A. S. (2012). Dyslexia in a second language? A dynamic test of reading acquisition may provide a fair answer. *Annals of Dyslexia, 62*, 172–185.
- Elliott, J. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review, 55*, 15–32.
- Feuerstein, R., Rand, Y., & Hoffman, M. B. (1981). The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments and techniques. *International Journal of Rehabilitation Research, 4*, 465–466.
- Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read. *Journal of Learning Disabilities, 44*, 339–347.
- Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78*, 263–279.
- 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.
- 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.
- Guthke, J. (1992). Learning tests—The concept, main research findings, problems and trends. *Learning and Individual Differences, 4*, 137–151.
- Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational applications*. New York: Cambridge University Press.
- Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired associate learning, phoneme awareness, and learning to read. *Journal of Experimental Child Psychology, 96*, 150–166.
- Jitendra, A. K., & Kameenui, E. J. (1993). Dynamic testing as a compensatory testing approach: A description and analysis. *Remedial and Special Education, 14*, 6–18.
- Litt, R. A., de Jong, P. F., van Bergen, E., & Nation, K. (2013). Dissociating crossmodal and verbal demands in paired associate learning (PAL): What drives the PAL–reading relationship? *Journal of Experimental Child Psychology, 115*, 137–149.
- O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3*, 159–197.
- Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading, 8*, 3–26.
- Petersen, D. B., Allen, M. M., & Spencer, T. D. (2014). Predicting reading difficulty in first grade using dynamic assessment of decoding in early kindergarten: A large-scale longitudinal study. *Journal of Learning Disabilities*. Advance online publication. doi:0022219414538518
- Seethaler, P. M., Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). Predicting first graders' development of calculation versus word-problem performance: The role of dynamic assessment. *Journal of Educational Psychology, 104*(1), 224.
- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*, 151–218.
- Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 72*, 95–129.
- Spector, J. E. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology, 84*, 353–363.
- Sternberg, R. J. (1996). *Successful intelligence*. New York: Simon & Schuster.
- Swanson, H. L. (2010). Does the dynamic testing of working memory predict growth in nonword fluency and vocabulary

- in children with reading disabilities? *Journal of Cognitive Education and Psychology*, 9, 139–165.
- Swanson, H. L., & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research*, 71, 321–363.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of word reading efficiency—Second edition (TOWRE-2)*. Austin, TX: PRO-ED.
- Tzuriel, D., & Haywood, H. C. (1992). The development of interactive dynamic approaches for assessment of learning potential. In H. Haywood & D. Tzuriel (Eds.), *Interactive assessment* (pp. 3–37). New York: Springer-Verlag.
- Vaughn, S., Denton, C. A., & Fletcher, J. M. (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychology in the Schools*, 47, 432–444.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Wagner, R. K., & Compton, D. L. (2011). Dynamic assessment and its implications for RTI models. *Journal of Learning Disabilities*, 44(4), 311–312.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing (CTOPP)*. Austin, TX: PRO-ED.
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, 33(3), 468–479.
- Warmington, M., & Hulme, C. (2012). Phoneme awareness, visual-verbal paired-associate learning, and rapid automatized naming as predictors of individual differences in reading ability. *Scientific Studies of Reading*, 16, 45–62.
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: Sage.
- Woodcock, R. W. (1998). *Woodcock reading mastery tests—Revised*. Circle Pines, MN: American Guidance Service.
- Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15, 642–656.