EXAMINING THE PREDICTIVE VALIDITY OF A DYNAMIC ASSESSMENT OF DECODING TO FORECAST RESPONSE TO TIER 2 INTERVENTION

By

Eunsoo Cho

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Special Education

May, 2014

Approved:

Donald L. Compton, Ph.D.

Lynn S. Fuchs, Ph.D.

TABLE OF CONTENTS

	Page
LIST OF TABLES	iii
Chapter	
I. INTRODUCTION	1
II. LITERATURE REVIEW	3
Measuring Responsiveness to Intervention: Curriculum-Based Measurement	3
Dynamic Assessment	4
III. METHOD	8
Participants and Procedures	8
Tier 2 Small-Group Tutoring	9
Measures	11
Data Analyses	14
IV. RESULTS	18
WIF Validation	20
Unconditional Model	21
Conditional Models	23
V. DISCUSSION	29
REFERENCES	36

LIST OF TABLES

Table		Page
1.	Student Demographics and Pre-Reading Level ……………………………..….……..	19
2.	Zero-order Correlation, Mean, and Standard Deviation for Growth Parameters and End-of-Year Reading Scores …………………………………..………...…..	20
3.	Zero-Order Correlation of Growth Parameters and Predictors ……………………………………………………………………………………….	22
4.	Unconditional Growth Models …………..	22
5.	Conditional Growth Models for Decoding Measures……………………………….	24
6.	Conditional Growth Models for Tier 1 WIF………………………………………...	26
7.	Conditional Growth Models for Pre-Reading Measures.…………………………...	28

CHAPTER I

INTRODUCTION

Response to Intervention (RTI) can be an alternative to the discrepancy model of identifying students with learning disabilities (LD; Fletcher et al., 2004; Fuchs, Mock, Morgan & Young, 2003; Gresham, 2002; Vaughn & Fuchs, 2003). Within an RTI framework, students may be identified with LD in part by their responsiveness to levels or tiers of increasingly intensive instruction. Students move from general instruction (Tier 1) to more intensive, explicit, and individualized instruction (Tier 2 and 3) according to their responsiveness to evidence-based instruction in prior levels. The success of an RTI system depends on correctly identifying students who are at elevated risk for poor academic outcomes as early as possible and placing them in a tiered instructional system. From an identification standpoint, a prescribed standard protocol consisting of evidence-based practices is typically provided in Tier 2. This instruction functions as a "test" for discriminating poor performers because of inappropriate instruction versus poor performers with intrinsic learning problems that limit response to generally effective instruction (Fuchs, Compton, Fuchs, Bryant, & Davis, 2008). Nonresponders to Tier 2 are considered at high risk of LD and receive more specialized instruction (i.e., Tier 3) as a result of unexpected failure to empirically proven instruction that is beneficial to a majority of students (Fuchs, Mock, Morgan, & Young, 2003). Thus, under a RTI model, the difference between false and true positives in LD identification is the "responsiveness" to Tier 2 evidence-based instruction.

Identification of Tier 2 nonresponders has become a critical LD classification issue. However, unresolved issues include how to conceptualize responsiveness to intervention and how this conceptualization can be operationalized to validly identify nonresponders (Fuchs & Deshler, 2007; Fuchs, Fuchs, & Compton, 2004). Researchers have used different ways to conceptualize and quantify responsiveness, including criterion- or norm-referenced posttreatment performance or progress monitoring results using curriculum-based measurements (CBM).

CHAPTER II

LITERATURE REVIEW

Measuring Responsiveness to Intervention: Curriculum-Based Measurement

CBM is an attractive measurement system for making data-driven decisions within a RTI framework because it permits practitioners to assess students on an ongoing basis and to identify those who show early signs of academic challenges (e.g., Fuchs, Fuchs, & Compton, 2004). The most widely accepted method of utilizing CBM in identifying nonresponders within a RTI is the "dual discrepancy" framework that captures both growth rate and current level of performance (Fuchs & Fuchs, 1998; Speece & Case, 2001). A dual discrepancy approach represents a more comprehensive understanding of responsiveness to instruction because it incorporates dynamic (growth) and static (performance level) indicators. Empirical evidence suggests one time measures of student performance are insufficient, and ongoing assessment to index growth may be necessary for valid identification of nonresponders to evidence-based reading instruction (Fuchs, Fuchs, & Compton, 2004; McMaster, Fuchs, Fuchs, & Compton, 2005; Speece & Case, 2001). Thus, both CBM growth and performance level may be needed to validly indicate responsiveness, especially for Tier 2 in which standard treatment protocols of evidence-based practices are delivered.

Even so, the provision of intervention, with CBM monitoring may unintentionally make RTI another wait-to-fail model (Compton et al., 2012). This is due to the time required to implement evidence-based practices to obtain reliable growth data. Another option is to directly measure students' responsiveness at one time point using dynamic assessment (DA).

Dynamic Assessment

DA refers to assessments that focus more on measuring the process of learning than on the final product of previously learned skills (Grigorenko & Sternberg, 1998). This is accomplished by providing scaffolds to students within a testing session. One of the most widely used scaffolding formats of DA is "graduated prompts." The tester provides a sequence of instructional tasks that increases in explicitness to help the student to succeed at the learning task. The number of prompts provided during the DA is an index of the student's responsiveness: the more prompts student needs, the less responsive he or she is. In a similar vein, the number of prompts indexes the degree of instructional explicitness required for a student to learn. Therefore, DA has potential as an RTI approach to LD identification. In fact, many have recognized the conceptual similarities between RTI and DA (e.g., Grigorenko, 2009; Wagner & Compton, 2011). They both merge intervention and assessment to close the gap between what is taught and tested. DA incorporates instruction into assessment and RTI incorporates assessment into instruction (Sternberg & Grigorenko, 2002).

Empirical evidence is accruing that DA has predictive validity for academic difficulties. Researchers have begun developing DAs specific to basic academic skills and testing their validity in forecasting later academic performance (Bridges & Catts, 2011; Compton et al., 2010; Fuchs, Compton, Fuchs, Bouton, & Caffrey, 2011; O'Connor & Jenkins, 1999). If the assessment content is directly linked to the academic domain and related curriculum, then results may be more sensitive to student learning and may help practitioners guide instruction (e.g., Haywood & Lidz, 2007; Campione & Brown, 1990). In the following, I focus the review on the studies of reading DA that predict word reading skills because (a) RTI models are typically implemented in early elementary grades to identify and remediate students at-risk for reading disabilities (RD); (b) most students with early reading problems have poor decoding and word reading skills (see Rack, Snowling, & Olson, 1992, for a review); and (c) Caffrey et al. (2008) reported varying results of predictive validity depending on the academic domain.

Some reading researchers developed DA in phonological processing such as segmenting (Spector, 1992; O'Connor & Jenkins, 1999; Zumeta, 2010) and deletion (Bridges & Catts, 2011) because of their discontent with the existing static measures, which show floor effects and unsatisfactory prediction accuracy for children who are at early risk for RD. Thus Spector (1999), Bridges and Catts (2011) and Zumeta (2010) used items in existing PA measures and incorporated graduated prompts to help students learn the tasks. O'Connor and Jenkins (1999) developed a segmentation DA with three levels of instruction to help students master segmenting novel words to onset and rime. Across these studies, DAs contributed 2% - 21% unique variance to later word reading of kindergarteners and first-grade students. Also, DA showed potential for improving prediction in identifying students with RD by reducing the false positives when used in conjunction with traditional phonological processing measures.

Another set of reading DA studies have employed decoding DA using graduated prompts to examine its construct and predictive validity. Caffrey (2006) compared CBM growth, CBM initial performance level, and DA in terms of respective predictive validity for end-of-year word reading in kindergarten and first-grade sample. For predicting word reading at the end of kindergarten, level of letter-sound knowledge fluency (a form of CBM) and DA were significant. For first graders, word identification fluency (WIF) growth and initial performance level along with DA were all significant. Results from commonality analysis indicated that DA explained the greatest amount of unique variance (13%) in later word reading followed by CBM growth (5%) and initial performance level (4%). Fuchs et al. (2011) examined the predictive validity of decoding DA in a representative sample of first-grade children comparing it against other well established early indicators of reading (e.g., rapid letter naming, oral vocabulary, phonological awareness etc.). Fuchs et al. found that DA was a significant predictor of word-level reading outcome, explaining 2.3% to 5.5% of unique variance. Decoding DA had less additive value when it was required to compete against all other pre-reading variables than compared to only CBM measures as in Caffrey's study.

Compton et al. (2010) proposed a more direct application of decoding DA in an RTI framework. They suggested a two-step gated screening procedure. In a first stage, a brief screener using a lenient cut-score is used to reduce true negatives to make the pool of screening students more manageable. The second stage focused on discriminating false positives from true positives using a multivariate screening battery. They added measures of first-grade WIF growth and DA in this second stage separately as a predictor of second-grade RD status. Adding WIF and decoding DA improved classification accuracy by decreasing the number of false positives. Also, WIF level and growth, and DA each uniquely contributed to prediction accuracy.

To summarize, the extant literature suggests DA of early reading skills has predictive validity in forecasting later word reading, explaining small but significant amounts of variance and enhancing classification accuracy. However, an important missing piece of information in the literature is whether DA can serve as a proxy for a child's potential to benefit from reading instruction by accurately predicting growth as well as final level of performance. All of the studies reviewed used either final level of word reading or RD status as a criterion but not growth. If DA indexes responsiveness as intended and if responsiveness is better captured by both performance level and growth, according to the dual discrepancy model, we need evidence of DA predicting growth in addition to final reading level. Although the outcome was not real word reading, Swanson (2010)'s study suggested the possibility of DA as a predictor of growth. Rather than using word reading tasks, he used cognitive tasks that underlie the reading process (verbal working memory) in DA, and he found students who needed more prompts to succeed with the tasks improved less in timed nonword reading tasks measured across three waves.

The research reviewed above provides evidence that the addition of DA in the prediction model can enhance prediction or classification accuracy of students with RD. Yet, there is still insufficient empirical evidence that DA is also predictive of growth or RTI. Hence, the present study investigated the role of DA in predicting responsiveness both in terms of final performance level and growth of students' word reading. More importantly, I situated this study in an RTI context so that students' responsiveness to a standardized, validated Tier 2 tutoring program was predicted. This could provide evidence for the utility of DA in a RTI decision-making process in predicting who will not respond to Tier 2 and thus need Tier 3.

In this study, I attempted to extend the research literature in three ways. First, whereas previous studies compared either timed or untimed static measures with DA (Bridges & Catts, 2011; Spector, 1992), I compared decoding DA's predictive validity against two standardized measures of decoding, timed and untimed. Second, I directly examined whether DA can help predict Tier 2 responsiveness, over the Tier 1 responsiveness measured by WIF, another dynamic indicator of responsiveness. Third, I built a prediction model for Tier 2 responsiveness by competing DA against important predictors of reading development (e.g., rapid letter naming, phonological awareness, oral vocabulary, and IQ). I examined whether DA has additional predictive power of responsiveness and whether its prediction ability is due to the shared common variance with language and IQ as suggested in Fuchs et al.'s study (2011) or whether it is a measure capable of capturing unique variance associated with responsiveness to Tier 2.

CHAPTER III

METHOD

Participants and Procedures

I used a sample from a larger evaluation of RTI efficacy (Compton, Fuchs, & Fuchs, 2006). In two cohorts, first-grade children who were identified as unresponsive to Tier 1 were involved in a randomized control trials examining the efficacy of 14 weeks of small-group Tier 2 intervention. Across the two cohorts, participant selection criteria and the intervention protocols were identical. The two cohorts were equivalent in their demographics and initial pre-reading measures, thus the data were combined for the present study (see Table 1).

In mid-September for each cohort, students nominated by their teachers as struggling readers were screened using three 1-min tests (i.e., two WIF lists and Rapid Letter Naming). Every nominated child who was consented were assessed (n = 624), and then a factor score was used to divide the 624 children into high-, average-, and low-performing groups (for details see Gilbert et al. in press). Children from the low group were retained for study. In this way, 438 children were identified as initially low performing. Beginning the first week of October, weekly WIF progress monitoring (PM) were administered for 6 weeks, each time with an alternate form while students received regular reading instruction in their classrooms (Tier 1). At the end of short-term PM, 10 (2.28%) of the children had moved from the district and were unavailable for assessment. One additional student was removed from the study due to scheduling difficulties.

Following short-term PM, individual growth modeling was used to select the students who were unresponsive to Tier 1. Because no agreed upon definition of response is available in the literature, unresponsiveness to Tier 1 was designated by rank ordering students on 6-week WIF intercept and slope then by selecting roughly the bottom half of the 427 students (232 students or 54.33% of the at-risk sample). Responders performed better on 6-week final level, $F(1,426) = 214.96$, $p < .0001$, $d = 1.99$ (responder $M = 27.83$, $SD = 8.86$; nonresponders $M = 13.65$, $SD = 5.22$), and growth, $F(1,426) = 116.77$, $p < .0001$, $d = 1.52$, per week (responder $M = 1.99$, $SD = 0.76$; nonresponders $M = 1.09$, $SD = 0.40$). In mid-November, the 232 students considered unresponsive to Tier 1 instruction received a battery of tests, administered individually by trained examiners (each of whom had demonstrated at least 90% accuracy during practice assessments). The battery comprised measures of phonemic awareness (PA), rapid letter naming (RLN), oral vocabulary (OV), DA, untimed word identification (WID), untimed word attack (WAT), sight word efficiency (SWE), and phonemic decoding efficiency (PDE). Among these students, four students did not assent to participate resulting in 228 students.

Approximately two-thirds of the nonresponders were randomly assigned to Tier 2 treatment (n=149) and others to control (n=79). Those who received Tier 2 small-group tutoring were used in this analysis to investigate the role of DA in predicting responsiveness to Tier 2. Finally, 15 students who had missing data on any of the predictor variables were dropped from the analysis, resulting in a final sample of 134.

Tier 2 Small-Group Tutoring

Tutoring

Tier 2 small group tutoring, supplementary to Tier 1, was provided in groups of three students for 14 weeks, three days a week, with each session lasting 45 mins. Treatment was considered Tier 2 because it comprised scripted, supplemental tutoring that focused on phonological awareness, sight words, letter sounds, decoding, and reading fluency (for additional information about tutoring see Gilbert et al., in press). Tutoring lessons were constructed around a series of eight leveled reading books and each session consisted of seven stages addressing an array of reading skills. Session begun with sight word reading followed by story words, letter-sound correspondence, phonological awareness and decoding (segmenting and blending sounds from letter-sound correspondence activities), spelling (spelling decodable word from the previous activity), sentence strip (locating the words read to them in the sentence), and passage reading (timed reading of a paragraph). For each activity, except spelling and sentence strips, tutors reviewed letter sounds and words from the previous session and introduced the new letter sounds and words for the day with a flashcard. When introducing new words or letters, the tutors modeled reading and spelling the words followed by students' choral and independent practice.

Approximately one third of the session was spent on word recognition in flashcards and in sentence strips. One third of the tutoring session was devoted to decoding instruction which involved letter-sound correspondence, segmenting and blending the decodable words, and spelling the words. For the remainder of the time, students worked on building reading fluency by reading the book. For each activity, if a student made mistakes, the tutor gave corrected feedback. Tutors gave points for good behavior and effort.

Tutors

Graduate students enrolled in the college of education were trained as research assistants (RAs) to implement Tier 2 small group tutoring. RAs participated in five weeks of tutor training with two hours of introduction to the tutoring scripts. Each RA completed 17 hrs of practice in the tutoring protocol, and small-group tutoring sessions were simulated with two other RAs and the trainer for fidelity. The trainer provided corrective feedback after the session.

Treatment fidelity

The RAs audiotaped every lesson, and a sample of 20% of the lessons was randomly selected for fidelity check. Implementation fidelity was monitored based on a fidelity checklist created from the tutoring script. Implementation fidelity reached an average of 94.04% of the steps.

Measures

Pre-intervention battery

Measures to predict individual differences in Tier 2 responsiveness were collected before Tier 2 intervention. These included rapid letter naming, phonemic awareness, oral vocabulary, IQ, untimed decoding, timed decoding, and dynamic assessment.

Rapid letter naming (RLN). The Comprehensive Test of Phonological Processing: Rapid Letter Naming (Wagner, Torgesen, & Rashotte, 1999) measures the speed at which an individual can name an array of 36 letters. The child names the letters as quickly as possible and the score is the number of sec required to complete the task. The test-retest reliability for the RLN subtest is .97 for children ages 5 to 7.

Phonemic awareness (PA). The Comprehensive Test of Phonological Processing: Sound Matching (SM; Wagner et al., 1999) assesses identification of first and last sounds in words, presented along with drawings depicting the words. To assess first sound matching, children are asked to determine which of three different words start with the same sound as the target word (e.g., “Which word starts with the same sound as ‘pan’? pig, hat, or cone?”). All words are presented as pictures to children. A parallel procedure assesses last sound matching. The test begins with three practice items and consists of 20 items. Split-half reliability exceeded .90 for the first-grade sample.

Oral vocabulary (OV). Woodcock-Johnson Psychoeducational Battery – Revised: Oral Vocabulary (Woodcock, McGrew, & Mather, 2001) assesses the ability to provide synonyms and antonyms in response to stimulus words presented orally. Split-half reliability exceeded .90 for the first-grade sample.

Intelligence (IQ). The two subtests of *Wechsler Abbreviated Scale of Intelligence* (WASI; The Psychological Corporation, 1999), Vocabulary and Matrix Reasoning, were used to measure full scale IQ. The Vocabulary subtest has 42 items that measure expressive vocabulary, verbal knowledge, and foundation of information. Students are asked to define or describe the meaning of the word presented orally by the examiner. Students' responses are scored from 0 to 2 depending on their quality. The test stopped if a student gave 5 consecutive zero responses. Matrix Reasoning measures nonverbal fluid reasoning with 35 items. Students are asked to find the correct picture among 5 choices that aligns with the series of three pictures in the pattern. The test stopped if a student gave 5 consecutive incorrect responses. Split-half reliability exceeded .85 for Vocabulary and .90 for Matrix Reasoning for the first grade sample.

Dynamic Assessment (DA). The same decoding DA that was used in the previous studies (i.e., Caffrey, 2006; Compton et al., 2010; Fuchs et al., 2011) was administered. Decoding DA is a scripted assessment for teaching and assessing reading skills of pseudowords with three decoding skills increasing in difficulty: CVC, CVCe (silent e), and CVC(C)ing (doubling). In CVC skill, students are taught to master the short 'o' vowel sound. In CVCe skill, students are taught to master the long 'o' vowel sound. In CVC(C)ing skill, students are taught the doubling rule and when the short and long 'o' sounds are used. For each skill, five levels of increasingly explicit instruction are embedded. Between each level of instruction, six pseudowords are presented to assess whether a student mastered the target decoding skill from the instruction. When a student reaches mastery criteria (five correctly read pseudowords out of six), he/she moves to the next decoding skill. If the student reads fewer than five words correctly, the tester moves to the next level of more explicit instruction. If the student fails to achieve mastery even after all five levels of scaffolded instruction, the session is terminated. The score is the number of scaffolding that the student needed to master the three decoding skills (0 = a student read at least five words correctly after the first instructional level; 5=a student did not reach mastery after all five instructional levels or, a student read at least five words correctly after the fifth instructional level). A four-week test-retest reliability was .72 in a pilot study (Fuchs, 2009; for additional information see D. Fuchs, Fuchs, Compton, Bouton, Caffrey, & Hill, 2007).

Untimed decoding. The Word Attack (WAT) subtest in the Woodcock Reading Mastery Test-R/NU (Woodcock, 1998) was used as a measure of untimed nonword reading in isolation. Students are asked to use grapheme-phoneme correspondence knowledge to read nonwords. The split-half reliability reported in the manual is .94 and inter-rater reliability exceeded .95 for the sample.

Timed decoding. The Test of Word Reading Efficiency: Phonemic Decoding Efficiency (TOWRE: PDE, Torgesen, Wagner, & Rashotte, 1997) was used to measure decoding accuracy and fluency. Students are asked to decode pseudowords accurately in 45 sec. Test-retest reliability was .86 for the first grade sample.

During intervention PM

WIF (L. S. Fuchs, Fuchs, & Compton, 2004) was used to monitor students' progress during the six-week screening period in Tier 1 and 14 weeks of small-group tutoring in Tier 2.

Students are presented with a single page of 50 high-frequency words randomly sampled from 100 high-frequency words from the Dolch pre-primer, primer, and first-grade level lists. Students were given one min to read the word list, and if they hesitated on a word for 3 sec, the examiner prompted them to proceed. The score for WIF is the number of words read correctly in 1 min. At weekly assessments, students were asked to read from 2 parallel WIF lists, and mean performance was calculated. Inter-rater reliability for WIF exceeded .89 (Zumeta, Compton, & Fuchs, 2012).

WIF validation measures

To validate WIF growth and final performance level used to indicate responsiveness to Tier 2 small group tutoring, I included four standardized word reading measures and one comprehension measure: The Word Identification (WID) and Word Attack (WAT) from the Woodcock Reading Mastery Test-R/NU for untimed reading, the Sight Word Reading Efficiency (SWE) and Phonemic Decoding Efficiency (PDE) from the Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1997) for timed reading, and passage comprehension from the Woodcock Reading Mastery Test-R/NU(Woodcock, 1998).

Data Analyses

Individual growth modeling (IGM) of WIF performance was used to examine the predictive validity of decoding DA in forecasting responsiveness during 14 weeks of Tier 2 small group tutoring. First, IGM was completed using HLM 6.0 (Bryke, Raudenbush, & Congdon, 1996) to estimate univariate growth parameters (i.e., final level and growth) of WIF during Tier 2 and to investigate how final performance level and growth in WIF differ across individuals. Because students were nested in classroom/tutoring group and schools, I first ran a 3-level unconditional means model, with time nested in child and child nested in school, to investigate whether there was dependency in the data at the school level. School level variance was marginally significant, 11.41, $\chi^2(10)=21.88$, p <.05, and intra-class correlation (ICC) for the school level was small (ρ=.06) (see Hedges & Hedberg, 2007). Because there was little school level variance to be modeled, I dropped it from the model. Next, I explored whether there were dependencies at the classroom or small-group level. Cross-classified models were used because students were neither nested in classrooms nor in small groups, but rather cross-classified to classrooms and small groups. For cross-classified models, intra-unit correlation coefficient (IUCC), which is equivalent to ICC for nested models, was calculated for the unconditional means model. IUCC describes the proportion of variance at a given level in relation to the total variance, calculated with the following equation: ρ

$$\rho_a = \frac{\tau_a}{\tau_a + \tau_b + \tau_c + \sigma^2}$$

where τ_a is the variance due to students, τ_b is the variance due to classroom, τ_c is the variance due to small group, and the σ^2 is the residual variance. IUCC for the full model revealed that only 1% of the variance was due to small groups and no variance was due to classrooms. I attribute this low dependency associated with small groups and classrooms on two facts. The use of standardized scripted protocol delivered by trained RAs likely limited small group variance, whereas the selection criteria of students who were unresponsive to general education instruction resulted in a lack of classroom effects. This suggests I ignore classroom/small group level dependency. The majority of the variance could be attributed to residual variance (i.e., variance at the level of the individual). In individual growth modeling, residual variance in unconditional grand-means model refers to the variation within person across time. This indicates the need for adding growth parameters; thus, growth parameters were included in the model.

To accurately capture WIF slope as a function of time, linear and quadratic growth parameters were fitted for the 17 assessment waves for WIF measures. Although tutoring was provided for 14 weeks, 17 WIF waves were used because three weeks were skipped due to regular school breaks (i.e., fall, winter, spring breaks). Each assessment wave was equally spaced (1 week). Week was coded from -17 to 0 for the linear slope term and from 289 to 0 for the quadratic slope. Time was centered at the end so intercept coefficient represents the mean value from the fitted model at the last week of tutoring (final level). The slope term represents the average amount of linear growth per week, and the quadratic slope term indicates the average curvature in growth rate per week. In terms of determining growth patterns, two models (i.e., linear growth with random intercept and linear slope vs. quadratic growth with random intercept, linear slope, and quadratic slope) were first compared based on the three criteria: randomness and homoscedasticity of the residuals, statistical significance of the fixed and random effects in quadratic growth model, and deviance statistics for model fit comparison.

Once the best unconditional growth model was identified, three separate series of conditional models were tested: decoding prediction model (WAT, PDE, and DA), Tier 1 responsiveness model (Tier 1 WIF growth, final level, and DA), and pre-reading model (RLN, PA, OV, IQ, and DA). First, to test the relative importance of DA over the static measures, WAT and PDE were introduced as level-2 individual characteristics, and DA was included in the second step. Second, to examine whether DA adds information beyond Tier 1 responsiveness based on WIF growth, individual's estimated growth and final level of WIF during 6 weeks of Tier 1 prior to entering Tier 2 were entered into the prediction model, and DA was added. Third, to test DA's superiority over early indicators of reading measures, I entered measures of RLN, PA, OV, and IQ as level-2 individual characteristics. Four steps of analysis were involved to answer the third, pre-reading, research question. First, RLN, PA, OV, and IQ were included as the predictors of final level and linear growth to answer whether these pre-reading skills can predict responsiveness to Tier 2 small-group tutoring. Second, nonsignificant predictors were removed from the model to yield a more parsimonious model. Third, DA was added as a predictor for final level and linear growth to determine if DA predicts growth after controlling for pre-reading skills. Fourth, nonsignificant predictors were deleted for the final model.

CHAPTER IV

RESULTS

Table 1 presents demographic information for two cohorts. No differences between the two cohorts were found. A series of comparisons using chi-square or *F* tests were conducted to detect whether any differences between cohorts existed in demographics, initial reading level, and cognitive characteristics. The two cohorts were initially equivalent except for Tier 1 slope, $F(1,132) = 4.22$, $p = .42$. Therefore, I combined cohorts for the remaining analyses. All participants were low readers who were eligible for Tier 2 because they did not show adequate response to Tier 1 compared to their peers. Results from pre-reading measures confirmed students' low reading levels (see Table 1). To provide a reference, I provide mean *z* scores for the Tier 2 sample, referenced against a representative sample selected from a developmental cohort (RLN = -.293, SM = -.534, OV = -.609, PDE = -.718, WAT = -.720, DA = -.638, WIF final level = -1.30, WIF linear growth = -.275).

Table 1

Student Demographics and Pre-Reading Level

Variables	Cohort 2 N (%)	Cohort 3 N (%)	Total N (%)	χ^2 (df)	p
Gender				3.13 (2)	.209
Male	31 (23)	36 (27)	67		
Female	36 (27)	29 (22)	65		
Unidentified	0 (0)	2 (1)	2		
Race				5.10 (3)	.165
African American	36 (27)	28 (21)	64		
Caucasian	26 (19)	26 (19)	52		
Hispanic	2 (1)	8 (6)	10		
Others	3 (2)	5 (4)	8		
ELL				1.01 (1)	.315
Non ELL	67 (50)	66 (49)	133		
ELL	0 (0)	1 (1)	1		
Free Lunch				.30 (1)	.583
No	24 (18)	21 (16)	45		
Yes	43 (32)	46 (34)	89		
IEP				.24(1)	.628
No	58 (43)	56 (42)	114		
Yes	9 (7)	11 (8)	20		
Retained				.75(1)	.784
No	59 (44)	60 (45)	119		
Yes	8 (6)	7 (5)	15		
	M (SD) N=67	M (SD) N=67	M (SD) N=134	F	p
Pre_WAT	5.30 (4.55)	6.67 (4.37)	6 (4.48)	F(1,132)=3.20	.076
Pre_PDE	4 (3.87)	4.76 (3.64)	4.38 (3.76)	F(1,132)=1.38	.243
Tier 1_Level	12.78 (4.91)	13.61 (5.23)	13.20 (5.07)	F(1,132)=0.91	.341
Tier 1_Growth	.85 (.69)	1.09 (.66)	.97 (.68)	F (1,132)=4.22	.042
Pre_RLN	66.16 (31.84)	67.67 (22.27)	66.92 (27.39)	F(1,132)=.10	0.751
Pre_SM	12.06 (5.12)	11.34 (5.02)	11.70 (5.07)	F(1,132)=.67	.415
Pre_OV	7.96 (3.81)	7.54 (4.26)	7.75 (4.03)	F(1,132)=.36	.550
Pre_IQ	89.67 (10.33)	91.16 (10.58)	90.42 (10.44)	F(1,132)=.68	.410
Pre_DA	11.67 (2.64)	11.25 (2.46)	11.46 (2.55)	F(1,132)=.90	.346

Note. ELL= English language learners; IEP= Individual Education Plan; Pre_RLN= pretest score of rapid letter naming; Pre_SM= pretest score of sound matching; Pre_OV= pretest score of oral vocabulary; Pre_IQ= pretest score of IQ; Pre_DA= pretest score of dynamic assessment.

WIF Validation

To validate whether WIF final level and growth parameter estimates are associated with students' reading level, zero-order correlations with end-of-first-grade reading performance were ran. Results are presented in Table 2. All correlation coefficients were statistically significant. WIF final level and growth showed similar patterns with end-of-first-grade reading measures although growth was less correlated with posttreatment reading scores than final level. They were highly correlated with timed and untimed word reading (r = .68-.89) and less highly correlated with timed and untimed word decoding (r = .45-.68) as well as passage comprehension (r = .59-.70). WIF final level was also highly correlated with their growth during tutoring (r =.89).

Table 2
Zero-order Correlation, Mean, and Standard Deviation for Growth Parameters and End-of-Year Reading Scores

	1	2	3	4	**5**	6	7
1. WIF_final level	-						
2. WIF_linear growth	.89	-					
3. Post_WID	.82	.68	-				
4. Post_WAT	.54	.45	.67	-			
5. Post_SWE	.89	.76	.88	.55	-		
6. Post_PDE	.68	.61	.73	.67	.70	-	
7. Post_PC	.70	.59	.82	.59	.76	.60	-
Mean	32.40	1.01	29.76	9.59	26.02	7.00	13.51
SD	16.83	.69	10.71	6.70	9.88	5.20	3.52
n	134	134	133[a]	133[a]	133[a]	133[a]	133[a]

Note. WIF_final level= the estimated performance of WIF in April; WIF_linear growth= the estimated rate of change of WIF per week during 14 weeks of tutoring; Post_WID= post test score of word identification; Post_WAT= post test score of word attack;Post_SWE= post test score of sight word efficiency; Post_PDE= post test score of phonemic decoding efficiency; Post_PC= post test score of passage comprehension
[a] One student had missing data on the posttest measures.

Unconditional Model

Zero-order correlation of growth parameters and predictors are presented in Table 3. Visual inspection of the residual plots suggested both models (i.e., linear and quadratic growth) produced random residuals. When the random quadratic term was added, the final level, linear growth, and quadratic growth parameters were all significant; yet, fixed and random effects of quadratic parameters were very small (see Table 4). Hypothesis testing with the deviance statistics suggest a quadratic growth model does fit better than the random linear model, $\chi^2_{(4)}$= 59.26, p < .000. Although both models are plausible models for explaining growth in the data, there was little variance to model in quadratic parameters (.002). I therefore selected the more parsimonious random linear slope model, because it lends itself to easier interpretations of the conditional models. Reliability for the individual growth parameters of final level (.981) and slope (.854) were high enough, indicating there were substantial signals in these data to be modeled. Results from the unconditional linear growth model with random final level and linear growth indicated students read approximately 32 words, $t(133) = 22.28$, $p < .0001$, at the end of Tier 2, and they gained approximately one word each week, $t(133) = 17.04$, $p < .0001$. By including random linear growth in the model, 85% of the variance from the grand-means model was explained, and there were still significant individual differences across students in both the final level and linear growth estimates.

Table 3

Zero-Order Correlation of Growth Parameters and Predictors

	1	2	3	4	5	6	7	8	9	10	11
1. WIF_final level	-										
2. WIF_linear growth	.89	-									
3. Pre_DA	-.55	-.41	-								
4. Pre_WAT	.50	.35	-.69	-							
5. Pre_PDE	.48	.35	-.58	.61	-						
6. Tier 1 final level	.64	.42	-.50	.44	.34	-					
7. Tier 1 linear growth	.27	.23	-.36	.24	.12	.43	-				
8. Pre_RLN	-.31	-.27	.20	-.15	-.20	-.33	-.20	-			
9. Pre_SM	.59	.49	-.56	.51	.43	.42	.08	-.20	-		
10. Pre_OV	.34	.23	-.49	.45	.41	.34	.13	-.02	.45	-	
11. Pre_IQ	.27	.18	-.32	.29	.28	.26	.10	-.08	.14	.34	-

Note. Coefficients above .17 are statistically significant at alpha level .05.

Table 4

Unconditional Growth Models

Linear Growth Model				Quadratic Growth Model			
Fixed effects	coefficient	SE	T (133)	Fixed effects	coefficient	SE	T (133)
Final Level	32.39***	1.45	22.28	Final Level	34.53***	1.49	23.15
Linear Growth	1.01***	0.06	17.04	Linear Growth	1.35***	.11	12.08
				Quadratic Growth	.02**	.01	3.23
Random effects	Variance	χ^2_{133}	Reliability	Random effects	Variance	χ^2_{133}	Reliability
Final Level	278.04***	7044.89	.981	Final Level	286.62***	3407.35	.959
Linear Growth	0.4***	913.57	.854	Linear Growth	0.90***	299.18	.541
				Quadratic Growth	.002	249.57	.458

Note. ****p* <.001. ** *p*<.01. *p<.05.

Conditional Models

DA over static decoding measures

In the first conditional model for decoding measures, two static decoding measures (i.e., WAT and PDE) were entered into the model to predict variance in the two growth parameters (final level and linear growth). Individual differences in WAT and PDE were unique predictors of WIF final level and linear growth during Tier 2. There was significant variance in final level, $\chi^2(131) = 4909$, $p < .000$, and linear growth, $\chi^2(131) = 780$, $p < .000$, after these static decoding measures were controlled. Reliability estimates indicated there was still substantial signal to be estimated for the final level (.974) and linear growth (.830). In the second phase of model building, decoding DA was added to the model. When static decoding measures were competed with decoding DA for final level and linear growth variance, DA was a significant predictor of both final level, $t(130) = -3.39$, $p < .05$, and linear growth, $t(130) = -2.48$, $p < .05$, controlling for the contribution of WAT and PDE. Final level variance uniquely explained by the decoding DA was 7% and linear growth was 6%. Although decoding skill was an important predictor of students' real word reading growth, the degree to which students need to master certain decoding rules had a unique contribution in explaining variance of final level and linear growth. It is important to note that WAT was no longer a significant predictor of final level, and both WAT and PDE lost their predictive power for explaining variances in linear growth. The decoding DA, which taps students' responsiveness, was in fact the only significant predictor of linear growth. Controlling for the current level of both timed and untimed decoding skills, students who need one more level of prompts for mastering the decoding rules read two words less and grew at a rate of 0.8 words per week. Model comparison of chi-square statistics suggested the second model with DA showed a better fit with the data, $\chi^2(2) = 12.43$, $p < .005$.

Table 5

Conditional Growth Models for Decoding Measures

Static Decoding Model				Full Decoding Model with DA			
Fixed effects	Coefficient	SE	T (131)	Fixed effects	Coefficient	SE	T (130)
Final Level							
Intercept	32.39***	1.23	26.39	Intercept	32.72***	1.18	27.61
WAT	1.21**	0.35	3.48	WAT	0.51	0.39	1.30
PDE	1.29**	.41	3.12	PDE	0.90*	0.41	2.19
				DA	-2.27**	0.67	-3.39
Linear Growth							
Intercept	1.01***	0.06	18.39	Intercept	1.02***	0.05	18.88
WAT	0.03*	0.02	2.20	WAT	0.01	0.02	0.61
PDE	0.04*	0.02	2.09	PDE	0.03	0.02	1.37
				DA	-0.8*	0.03	-2.48
Random effects	Variance	χ^2_{131}	Reliability	Random effects	Variance	χ^2_{131}	Reliability
Final Level	196.02***	4909.07	.974	Final Level	181.55***	4533.17	.971
Linear Growth	0.34***	780.79	.830	Linear Growth	0.32***	745.47	.823

Note. * p <.05. ** p < .01. *** p < .001.

DA over Tier 1 responsiveness measures.

Next I examined whether decoding DA has predictive validity in addition to six weeks of Tier 1 responsiveness indicators (i.e., Tier 1 final level and linear growth). In the first conditional model, WIF final level and linear growth (based on six weeks of PM in Tier 1) were entered into the prediction model. Results of the fixed effects indicated variance in Tier 2 final WIF level was explained by individual differences in Tier 1 final level but not by Tier 1 linear growth. Students who read one more word than other students at the end of Tier 1 read two more words at the end of Tier 2 and grew at a .05 words faster each week. The linear growth of WIF during Tier 1 did not explain Tier 2 responsiveness. There was significant variance in final level, χ^2 (131) = 4193, $p < .000$, and linear growth, χ^2 (131) =755, $p < .000$, after these Tier 1 responsiveness indicators were controlled for. Reliability estimates indicate there was still substantial signal to be estimated for the final level (.969) and linear growth (.824). This allowed further examination of whether decoding DA could explain some of the unexplained variance in the first analysis. Results of the second phase showed similar pattern with regards to Tier 1 responsiveness. When decoding DA was included in the model, it significantly predicted Tier 2 growth parameters above and beyond Tier 1 growth parameters, $t(130) = -4.44$, $p < .001$, for final level, $t(130) = -2.97$, $p < .001$, for slope. Final level variance uniquely explained by the decoding DA was 13% and linear growth was 6%. Tier 1 linear growth remained nonsignificant but Tier 1 final level remained significant for predicting Tier 2 final level, $t(130) = 6.61$, $p < .001$, and linear growth, $t(130) = 3.05$, $p < .001$, even after decoding DA was controlled. Controlling for Tier 1 responsiveness, students who needed one more level of prompts for mastering decoding rule read two words less and grew at a rate of .94 words each week. Model comparison suggested the second model with DA was a better fit with the data. χ^2 (2) =22.90, $p < .001$.

Table 6

Conditional Growth Models for Tier 1 WIF

Tier 1 WIF Model				Tier 1 Model with DA			
Fixed effects	Coefficient	SE	T (131)	Fixed effects	Coefficient	SE	T (130)
Final Level							
Intercept	32.40***	1.13	28.72	Intercept	32.40***	1.05	30.71
Tier 1 final level	2.12***	0.25	8.59	Tier 1 final level	1.67***	0.25	6.61
Tier 1 linear growth	-0.03	1.83	-0.017	Tier 1 linear growth	-1.54	1.74	-0.88
				DA	-2.16***	0.49	-4.44
Linear Growth							
Intercept	1.01***	0.05	18.71	Intercept	1.01	0.05	19.26
Tier 1 final level	0.05***	0.01	4.51	Tier 1 final level	0.04**	0.01	3.05
Tier 1 linear growth	0.06	0.09	0.68	Tier 1 linear growth	0.01	0.09	0.12
				DA	-0.07**	0.02	-2.97
Random effects	Variance	χ^2_{131}	Reliability	Random effects	Variance	χ^2_{131}	Reliability
Intercept	165.19***	4193	.969	Final Level	143.84***	3619	0.964
Linear Growth	0.32***	755	.824	Linear Growth	0.30***	705	0.814

Note. * p <.05. ** p < .01. *** p < .001.

DA over pre-reading measures

Lastly, I examined whether decoding DA has predictive validity over the early indicators of reading development. The hypothesized correlates of change in WIF during Tier 2 (i.e., RLN, SM, OV, and IQ), were first included in the prediction model. The proportion of the variance explained in the final level by these predictors was 41%; in the linear growth, 30%. Individual differences in RLN, PA, and IQ explained the variance in final level; only RLN and PA explained the variance of linear growth. After these pre-reading measures were partialled out there was still significance variance in the final level, $\chi^2(129)=4051$, $p < .001$, and linear growth, $\chi^2(129)=660$, $p < .001$, with high reliabilities (.968 and .803, respectively). The second step involved removing the nonsignificant predictors from the model to build the most parsimonious prediction model and adding decoding DA to the model. Results indicated DA was a significant predictor of Tier 2 responsiveness in the presence of other significant pre-reading measures. DA predicted final level, $t(130) = -3.51$, $p <. 01$, and also linear growth, $t(130) = -1.88$, $p < .001$. Others being equal, students who need one more level of prompts for mastering decoding rule read two words less and grew at a rate of one word each week. Final level variance uniquely explained by the decoding DA was 7% and linear growth was 3%. All other predictors remained significant. Model comparison of chi-square statistics suggested the second model with DA showed better fit to the data, $\chi^2(2) = 15.07$, $p < .05$.

Table 7

Conditional Growth Models for Pre-Reading Measures

Pre-reading Model				Pre-reading Model with DA			
Fixed effects	Coefficient	SE	T (129)	Fixed effects	Coefficient	SE	T[a] (129/130)
Final Level							
Intercept	32.868***	1.124	29.23	Intercept	32.94***	1.082	29.94
RLN	-.120**	.042	-2.83	RLN	-.106*	.040	-2.60
PA	1.664***	.255	6.52	PA	1.286***	.261	4.92
OV	.232	.331	.70	IQ	.117*	.057	2.03
IQ	.273*	.115	2.37	DA	-1.842**	.524	-3.51
Linear Growth							
Intercept	1.026***	.051	20.07	Intercept	1.011***	.050	20.36
RLN	-.004*	.002	-2.16	RLN	-.004*	.002	-2.61
PA	.060***	.011	5.22	PA	.048***	.012	3.75
OV	-.001	.015	-.12	DA	-0.049*	.024	-1.88
IQ	.007	.005	1.34				
Random effects	Variance	χ^2_{129}	Reliability		Variance	$\chi^2_{129/130}$a	Reliability
Intercept	163.60***	4051.50	.968	Final Level	151.55***	3783.94	.966
Linear Growth	.28***	660.16	.803	Linear Growth	.27***	648.62	.798

Note. [a] Degrees of freedom for the final intercept is 129 and slope is 130 for the DA model.

* p <.05. ** p < .01. *** p < .001.

CHAPTER V

DISCUSSION

I investigated the predictive validity of decoding DA for Tier 2 responsiveness beyond static decoding measures, Tier 1 responsiveness, and early indicators of reading development (RLN, PA, OV, and IQ). In the earlier prediction studies on DA, researchers used end-of-tutoring performance on standardized measures, pre-post difference scores, or LD designation as the criterion to be predicted. While these approaches provide some evidence of the role of DA in predicting future reading performance or classifying LD, they do not provide direct evidence of the predictive validity of DA for forecasting students' response to instruction (i.e., growth).

Based on the prior validity evidence of the dual-discrepancy method, I considered both final performance level and growth on CBM as important indicators of responsiveness (Fuchs & Fuchs, 1998). The present research extends the DA literature by evaluating the predictive validity of decoding DA using a comprehensive indicator of responsiveness, final level or growth, as the criteria for the prediction. Results are surprisingly consistent across three sets of individual growth models, that is, decoding DA significantly explained small but unique variance in Tier 2 responsiveness, both final level of WIF and growth. Although I did not simultaneously predicted final level and growth as suggested by the dual discrepancy method, the results lead me to conclude that DA may be a useful tool for improving prediction accuracy for identifying responders and nonresponders to Tier 2 beyond that accounted for by the static decoding measures, Tier 1 responsive measures, and precursors of reading development.

There has been a concern in the field that Tier 2 may not be necessary for struggling readers who will be nonresponders to Tier 2. For these small groups of students, they receive Tier 2 just to show their failure to receive Tier 3 intensive instruction (Compton et al., 2012; L.S. Fuchs, Fuchs, & Compton, 2010; Vaughn, Denton, & Fletcher, 2010). Compton et al. (2012) recently showed that adding data gained from Tier 2 (Tier 2 level, slope, and tutor ratings) to universal screening, Tier 1 data, and norm-referenced test did not improve classification accuracy of Tier 2 nonresponders. The authors concluded that Tier 2 data may not be necessary to identify students for Tier 3. If we can predict students' response to Tier 2 with increased accuracy via adding DA in a prediction battery, we may be able to accelerate the RTI process for nonresponders and provide Tier 3 intervention in a timely manner. However, it is also important to note that given the small amount of variance in Tier 2 responsiveness uniquely explained by DA (3%-13%), the conclusion made from this study is only suggestive. There are no set criteria in the literature about how much variance should be explained in outcome for a measure to have a practical significance, because it will depend on the type and number of predictors included in the model. Although the present result is consistent with the ones reported in other decoding DA studies (i.e., Caffrey, 2006; Fuchs et al, 2011) reporting 5.6 %-13% of unique variance explained, more work needs to be done to determine the practical significance of DA. One example of such work would be examining DA's classification accuracy (e.g., sensitivity of 90% and specificity of 80%) when added to other diagnostic measures (e.g., Compton et al., 2010). Another example is to investigate the social validity of DA. Although people might find the idea of DA appealing, some have stated that DA is not useful because of the amount of training required to administer and because of time constraints (Grigorenko, 2009). Thus, efforts to make DA administration more user-friendly should be made for DA to have a practical utility.

I now turn my attention to interpreting the results of each model. In the first prediction model, I compared decoding DA against the two standardized measures of decoding. By including both timed and untimed measures of decoding, I intended to capture the static aspects of decoding more comprehensively than the previous studies that competed DA against only timed (Bridges & Catts, 2011) or untimed (Bridges & Catts, 2011; Fuchs et al., 2011; Spector, 1992) measures. In explaining Tier 2 final WIF level, untimed decoding was not significant beyond timed decoding and decoding DA, although a moderate bivariate correlation (.50) was found. On the other hand, timed decoding which was also moderately correlated with the final level (.48) remained significant after controlling for the variance associated with untimed decoding and decoding DA. One possible explanation as to why untimed decoding lost its predictive power, but not timed decoding, with regards to the final level is because the outcome (WIF) was a timed word reading measure. In terms of Tier 2 linear growth, timed decoding lost its predictive power in the presence of decoding DA. Decoding is a primary tool for orthographic word reading (Share & Stanovich, 1995) and a proxy for word reading development (Compton, 2000). The present results, however, indicated measuring the process of learning decoding rule via DA, rather than assessing already developed decoding skill, was the better correlate of growth in real word reading during Tier 2 tutoring.

The second prediction model compared the two methods of indexing responsiveness: CBM and DA. I simultaneously entered Tier 1final WIF level and linear growth in the prediction model and then added DA. Results suggested that linear growth during Tier 1 did not predict their response to Tier 2 tutoring. I discuss three possible statistical explanations. First, the reason why linear growth of WIF during Tier 1 could not predict Tier 2 responsiveness may be the phenomenon of “bouncing betas” due to multicollinearity between final level and linear growth estimates. When two or more predictors are highly correlated, there is little unique information to be estimated resulting in instable coefficients. Previous studies on first graders' word reading growth indicate final word reading level and linear growth are highly correlated (Compton, 2000). However, this hypothesis is not supported because Tier 1 final level and slope had only moderate correlation (.43).

Second, when the variance is truncated, reduced correlation is observed compared to the unrestricted population. The participants included in this study were low readers, who were screened based on their final WIF level and linear growth during Tier 1. Clearly, when compared with the representative sample, the standard deviation for Tier 1 final WIF level decreased from 26.32 to 5.07; for WIF linear growth from 1.74 to .68. Nevertheless, Tier 1 final level still predicted Tier 2 responsiveness, suggesting that range restriction was not the primary reason for the lack of predictive validity for Tier 1 slope.

Third, low reliability attenuates correlation. Reliability of Tier 1 final level was .88; linear growth was .20. Reliability in growth modeling indicates the proportion of the observed variance to the true variance. Thus, low reliability of linear growth may suggest that small portion of the individual differences will be the true differences in the population. Alternatively, it may suggest a small amount of variability in linear growth exits in the population (Singer & Willett, 2003). Indeed, there was little observed variance of linear growth (.09) which might result in low reliability of the linear growth estimates. Then, does this mean Tier 1 linear growth is not useful for predicting who will respond to Tier 2? Or do we need to extend progress monitoring in Tier 1 to acquire reliable estimates of liner growth? There is no clear-cut answer to these questions as the literature has been inconsistent on the predictive validity of growth (i.e., slope) with varying participant characteristics, CBM sampling procedures, and criterion measures (Compton et al., 2010; Schatschneider, Wagner, & Crawford, 2008).

Now I turn to a more substantive account for the failure of Tier 1 linear growth's predictive ability for Tier 2 responsiveness. One may speculate about the qualitative difference between Tier 1 classroom instruction and Tier 2. The nature of Tier 1 general education classrooms is more difficult to control than Tier 2 instruction, which relied on a standardized tutoring program in the present study. Although RTI models posit that students receive high quality core instruction in Tier 1, this is not guaranteed. Therefore, responsiveness to Tier 1 and Tier 2 may show different patterns because responsiveness is a function of student characteristics as well as the nature of instruction. Because I did not collect data for the quality/fidelity of Tier 1, I could not explore this possibility in the present study. Although inexplicable issues remain, the clear conclusion is that decoding DA was superior to Tier 1 linear growth in explaining individual differences in Tier 2 responsiveness.

The third prediction model tested whether DA was an independent predictor of responsiveness above and beyond important precursor skills of reading (RLN, PA, OV, and IQ). A previous exploratory factor analysis indicated that DA loaded on a factor represented by IQ and language (Fuchs et al, 2011). This suggests that DA may not independently predict reading growth in a model that also includes language measures and IQ. Toward this end, I built a parsimonious prediction model including RLN, PA, OV, and IQ as predictors and then entered DA.

RLN and PA have both been shown to be important predictors of responsiveness in the literature (Al Otaiba & Fuchs, 2002; Nelson, Benner, & Gonzalez, 2003). They both explained significant proportions of variance in the final WIF level and linear growth during Tier 2 tutoring, even after controlling for language (oral vocabulary), IQ and decoding DA. However, the role of IQ in LD identification has been contentious in the field. Al Otaiba and Fuchs (2002) reported inconsistent findings across studies on whether IQ is related to student responsiveness to instruction. Nelson, Benner, and Gonzalez (2003) extended Al Otaiba and Fuchs' study using meta-analysis and found a significant relationship between IQ and responsiveness to Tier 2 instruction. The present results suggest that individual differences in IQ explain variance in final level of word reading skills but lacks predictive validity in predicting linear growth. In terms of oral language, previous studies have shown that vocabulary is a significant predictor of responsiveness and later word recognition (Al Otaiba & Fuchs, 2006; Catts, Fey, Zhang, & Tomblin, 1999). However, oral vocabulary was not a significant predictor of Tier 2 responsiveness in the present study for final level and linear growth. One possibility for this inconsistency is the differences in the type of vocabulary measured. Previous studies used receptive or a combination of receptive and expressive vocabulary measure. I indexed vocabulary, by contrast, with only one measure of expressive vocabulary. This difference may have resulted in different patterns from previous studies. Another possible reason for oral vocabulary's lack of predictive power is the nature of the Tier 2 instruction. The small-group Tier 2 tutoring did not address vocabulary or comprehension, rather it focused on word reading.

When added to the reduced prediction model in which RLN, PA, and IQ were retained as predictors of Tier 2 final level, and RLN and PA retained as predictors of Tier 2 linear growth, decoding DA was a significant predictor of both Tier 2 final level and linear growth controlling for well established pre-reading measures. This not only suggests the possibility of using DA as a useful supplemental measure for identifying Tier 2 nonresponders but also provides evidence of the construct validity of decoding DA as a measure of responsiveness.

Finally, results of the present study should be interpreted within the contextual framework of the study. First, participants were poor readers who already had demonstrated poor progress in response to general education instruction. Second, the criterion of responsiveness, WIF final level and linear growth during Tier 2, focuses on word-level reading skills. The predictive validity of DA in forecasting responsiveness may not hold if a different criterion were used. Third, Tier 2 was a standardized scripted protocol of instruction implemented with high fidelity. The results may not hold for RTI models with a problem solving approach.

In addition, the present study leaves some questions unanswered and future research should extend this work by exploring the construct validity of DA with a confirmatory approach. For example, using a structural equation modeling approach we can derive a latent construct of "responsiveness" using multiple indicators of responsiveness. Moreover, it is possible to test whether decoding and responsiveness represent a unitary or distinct construct using static and decoding measures of the same construct, decoding. Another lingering question is whether the emphasis of instruction moderates DA's predictive validity. In existing studies, differences in classroom instruction were conceived as a nuisance variable to be controlled by explaining the similarities in classroom instruction (Bridges & Catts, 2011) or statistically controlling (Fuchs et al., 2011). Researchers rarely incorporate the types or characteristics of classroom instruction students receive in Tier 1 or Tier 2 into the study design. Using multi-level modeling techniques, we can model this variance and test whether the type of classroom instruction moderates DA's predictive power. This will advance the understanding of how and why DA can be used for indexing Tier 2 responsiveness.

REFERENCES

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

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

Bridges, M. S. (2009). *The use of a dynamic screening of phonological awareness to predict reading risk for kindergarten students.* Unpublished doctoral dissertation, University of Kansas, Lawrence.

Bridges, M. S., & Catts, H. W. (2008). *Dynamic screening of phonological awareness (Prepublication version).* Moline, IL: Linguisystems.

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.

Caffrey, E. (2006). *A comparison of dynamic assessment and progress monitoring in the prediction of reading achievement for students in kindergarten and first grade.* Unpublished doctoral dissertation, Vanderbilt University, Nashville.

Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *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). New Jersey: Lawrence Erlbaum.

Catts, H. W., Fey, M. E., Zhang, X., Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3,* 331-361.

Compton, D. L. (2000). Modeling the growth of decoding skills in first grade children. *Scientific Studies of Reading, 4,* 219-259.

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 interventions: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102,* 327-341.

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., Barquero, L. A., & 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.

Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37,* 184-192.

Fletcher, J. M, Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia, 54,* 304-331.

Francis, D. J., Schatschneider, C., & Carlson, C. D. (2000). Introduction to individual growth curve analysis. In D. Drotar (Ed.), *Handbook of research in pediatric and clinical child psychology* (pp. 51-73). New York: Kluwer Academic.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 93-99.

Fuchs, D. (2009). *Dynamic assessment and RTI.* Coronado, CA: Pacific Coast Research Conference.

Fuchs, D., & Deshler, D.D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice, 22*, 129-136.

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

Fuchs, D., Deshler, D. D., & Reschly, D. J. (2004). National research center on learning disabilities: Multimethod studies of identification and classification issues. *Learning Disability Quarterly, 27,* 189-195.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-Intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice, 18,* 157-171.

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

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2010). Rethinking response to intervention at middle school and high school. *School Psychology Review,* 39(1), 22-28.

Fuchs, L.S., & Fuchs, D. (1998). Treatment validity: A unifying concept for the identification of learning disabilities. *Learning Disabilities Research and Practice, 14*, 204-219.

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(5),* 58-63.

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

Gresham, F. M. (2002). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D.P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 467- 519). Mahwah, NJ: Erlbaum.

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.

Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice.* Cambridge: Cambridge University Press.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlations for planning group randomized experiments in rural education. *Journal of Research in Rural Education, 22,* 1-15.

Jenkins, J. R. (2003, December). Candidate measures for screening at-risk students. Paper presented at NRCLD Responsiveness-to-Intervention Symposium, Kansas City, MO. Retrieved from http://www.nrcld.org/symposium2003/jenkins/index.html

Nelson, J. R., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research and Practice, 18,* 225-267.

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

O'Connor, R. E., & Jenkins, J. R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3,* 159-197.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27,* 29-53.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Application and data analysis methods* (2nd Ed.). Thousand Oaks, CA: Sage.

Schatschneider, C., Wagner, R. K., & Crawford, E. C. (2008). The importance of measuring growth in response to intervention models: Testing a core assumption. *Learning and Individual Differences,* 18, 308–315.

Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: A model of acquisition and individual differences. *Issues in education: Contributions from educational psychology, 1*, 1-57.

Singer, J. D., & Willett, J. B. (2003) *Applied longitudinal data analysis: Modeling change and event occurrence.* London: Oxford Univ.

Spector, J. E. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology, 84,* 353-363.

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

Sternberg, R. J., & Grigorenko, E. (2002). *Dynamic Testing: The Nature and Measurement of Learning Potential*. Cambridge: Cambridge University Press.

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.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). *Test of word reading efficiency.* Austin, TX: Pro-Ed.

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.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to intervention: The promise and potential problems. *Learning Disabilities Research and Practice, 18*, 137-146.

Wagner, R. K., & Compton, D. L. (2011). Dynamic assessment and its implication for RTI models. *Journal of Learning Disabilities*, *44,* 311-312.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A. (1999). *Comprehensive test of phonological processing.* Austin, TX: PRO-Ed.

Wechsler, D. (1999) *Wechsler abbreviated scale of intelligence.* San Antonio, TX: Psychological Corporation.

Woodcock, R.W. (1998). *Woodcock reading mastery test – Revised/Normative Update.* Circle Pines, MN: American Guidance Service.

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Zumeta, R. O. (2010). *Enhancing the accuracy of kindergarten screening.* Unpublished doctoral dissertation, Vanderbilt University, Nashville.

Zumeta, R. O., Compton, D. L., & Fuchs, L. S. (2012). Using word identification fluency to assess first-grade reading development: A comparison of two word-sampling approaches. *Exceptional Children, 78,* 201-220.