CONSTRUCT AND INCREMENTAL VALIDITY OF DYNAMIC ASSESSMENT OF DECODING WITHIN AND ACROSS DOMAINS

By

Eunsoo Cho

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

May, 2014

Approved:

Donald L. Compton, Ph.D.

Douglas Fuchs, Ph.D.

Lynn S. Fuchs, Ph.D.

Sonya K. Sterba, Ph.D.

Jennifer K. Gilbert, Ph.D.

To my wonderful parents for their unfailing support,

my amazing husband Eunseok Choi for his boundless love,

and to my beloved son Young-joon Choi

ACKNOWLEDGEMENTS

I owe a tremendous amount to many great people I have met throughout my journey in graduate school. A special thank you goes to my advisor, Dr. Don Compton, who took a risk when he invited an unknown international student to study special education at Vanderbilt. He has always encouraged me in times of frustration, guided me with thoughtful feedback, and inspired me with his deep insights. I am also grateful to Drs. Doug and Lynn Fuchs for their counsel and encouragement. Their support has helped me to develop confidence as a young researcher. My research was sparked by their work and their feedback has helped me distill my ideas. I also thank Dr. Sonya Sterba for being so generous with her time and expertise. I owe a great deal to my dear friend, mentor, and committee member, Dr. Jenny Gilbert. I cannot thank her enough for always being supportive on many different levels; from prayers to academic guidance throughout all stages of my graduate career. I thank the Semmel Family and the Dunn Family for their generous financial contributions, which have supported my graduate studies.

In addition, I am thankful for many others who are now my proud colleagues. I thank my caring cohort, the Significant Seven, for making my years at Vanderbilt indeed significant. I also thank my colleague, Laura Steacy, who has shared all the joys and despairs that we all experience as doctoral students. I am appreciative to Alyson Collins, Esther Lindstrom, Amanda Miller, Berit Barr, and Erika Hsu for helping me collect my dissertation data. This research would not have been possible without their help.

I am immensely grateful to all of my family members. I thank my dad who taught me the lesson that it is the effort that counts, my mom who has supported me in all the ways possible, and my sister who always encouraged me not to settle for less. I am blessed to have met my

amazing husband, Eunseok. It was his boundless support and unconditional love that made this accomplishment possible. During graduate school, I have also been blessed with my son, Young-joon. I hope my journey can inspire him to follow his passion, to not limit himself, and to positively contribute to the lives of others. Last but not least, I am thankful to the Lord Jesus for sharing His vision, leading me down this path, and being always faithful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

A conventional way to understand an individual's cognitive abilities is to assess independent performance as a final learned product; thus, traditional assessment is suited for phenomena that are static in nature. However, this static assessment may not inform what one will be able to do in the near future nor does it often reflect one's process of learning, but only what has already been learned (e.g., Sternberg, 1996; Tzuriel & Haywood, 1992). An alternative to static assessment is dynamic assessment (DA). DA is an umbrella term for assessment procedures that embed interaction between the examiner and examinee within the test (Lidz, 1987; Grigorenko & Sternberg, 1998). This allows the examiner to make judgments about students' responsiveness to instruction. The students' responsiveness then serves as a measure of learning potential, which is less dependent on prior educational history or preexisting knowledge and skills than static assessment. Because DA was developed to measure learning potential, a distinct construct from what is measured in static assessments, it is believed to provide additional information about academic achievement beyond what can be gathered from static assessments alone: hence, DA may be diagnostically useful in educational settings (for a comprehensive DA review, see Grigorenko & Sternberg, 1998).

Measuring learning potential is intuitively appealing and some empirical evidence suggests that DA has unique, but small, additive value in the prediction of concurrent and future academic achievement (see Caffrey, Fuchs, & Fuchs, 2008). However, further work is needed to ascertain whether this new construct of learning potential, as measured with DA, is valid and

superior to other means of measuring the related constructs in explaining academic achievements. The purpose of this study, therefore, was to examine the nature of learning potential that is measured via DA of decoding by addressing the following three questions: (a) Is DA measuring learning potential for acquiring decoding skills distinct from conventional static measures of IQ and decoding?; (b) Does DA of decoding have incremental validity in explaining word reading skills beyond that which can be explained by the known predictors of early reading development?; and (c) Does DA of decoding have incremental validity across domains in that it can explain arithmetic performance beyond the predictors of math?

CHAPTER II

LITERATURE REVIEW

Static Assessment versus Dynamic Assessment

Conventional ways of assessing cognitive abilities use static methods to measure pre-existing knowledge and skills. With these methods, no feedback is given during the testing session, and examiners stay in a neutral relationship with examinees. Furthermore, examiners adhere to the standardized testing procedure, and any deviance from it is considered a threat to the validity of the test score interpretation. The initial purpose of the development of such static intelligence tests was to select persons for military, educational, and industrial placement so that outcome could be maximized given the least investment (Lidz, 1987). In contrast, one of the unique purposes of educational assessment is to forecast academic achievement in order to select students who may need special instructional assistance so that instruction can be modified, allowing them to achieve their full potential. Critics of static assessments (including intelligence tests) assert that such tests do not provide the information needed to meet the goals of the educational assessment. These critics believe that static assessments fail to provide information about intra-individual change, neglect to sample enough items of basic skills which precludes sensitivity in identifying low performing students (i.e., floor effects), and are unable to discriminate poor performances due to intrinsic cognitive deficits from those with lack of educational opportunities (Sternberg, 1996). In response to these failures, DA has gained attention as an alternative (Grigorenko & Sternberg, 1998; Jitendra & Kameenui, 1993).

DA emphasizes assessment of the potential for learning rather than past learning. Therefore, assistance is given during the test either in the form of instruction or in a sequence of progressively explicit prompts. Examiners no longer stay neutral, but interact with examinees. In this way, DA indexes the individual's learning potential either by quantifying gains during the assisted phase or by indexing the number of prompts required to master the given task.

The theoretical root of DA can be traced back to the work of Lev Vygotsky (1934/1962) and his notion of the zone of proximal development (ZPD). According to Vygotsky, children's cognitive abilities can be fully understood by recognizing the two developmental levels: the *actualized* and the *actualizing*. The actualized abilities are seen as those that are complete and fully developed, reflecting what students have learned, whereas the 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. Consider two students who perform similarly poorly on a conventional static assessment. One may require little assistance to reach the desired competence level, whereas the other may need extensive support. In this example, the former is considered to have a higher learning potential than the latter. Thus, the ZPD is considered an important source of individual differences in learning, and DA was developed in an attempt to measure ZPD.

The construct being measured via DA is theoretically different from those assessed by static measures. DA focuses on future performance, that is, individuals' ability to respond to instructions (i.e., learning potential), whereas static measures only focus on independent performance at the time of the testing (i.e., what has already been learned). In other words, DA estimates the upper boundary of ZPD, which is how well an individual can learn given assistance, whereas static assessments measure the lower boundary of ZPD, which is what has been learned

and how well he/she can perform independently. DA provides additional information to that provided by static assessments by comprehensively explaining individual differences in learning.

Currently, there is no consensus about how best to assess students' learning potential. In fact, several different approaches to DA exist including learning potential assessment (e.g., Budoff, 1967), testing-the-limits (Carlson & Wiedl, 1979), mediated assessment (e.g., Feuerstein, 1979), and assisted learning and transfer (e.g., Bransford, Delclos, Vye, Burns, & Hasselbring, 1987; Campione, Brown, Ferrara, Jones, & Steinberg, 1985). These approaches differ in terms of the nature of the interaction (e.g., standardized vs. individualized), the format (e.g., test-teach-retest vs. graduated prompts), and how 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 typically incorporates individualized interaction with a blocked scheduling of instruction between pre- and posttest to index the improvement on the posttest. Alternatively, the graduated prompts approach uses progressive scheduling of a predetermined hierarchy of prompts and assesses the amount of help students require to master the skill. With the ease associated with using standardized prompts, the graduated prompts approach has been widely accepted by researchers in school settings interested in academic achievement for screening and identification of students with special needs (Daniel, 1997). Hence, the present study uses the graduated prompts DA, which provides a predetermined set of increasingly helpful prompts.

Two Challenges in Dynamic Assessment Research

When new constructs are proposed and assessment tools are developed, two things must be accomplished (Lubinski, 2004). First, the construct has to be validated within a nomothetic span that establishes the network of correlates surrounding measures of the construct to see how

it is different from and similar to existing measures (Embretson, 1983). Second, incremental validity, which refers to the degree to which a measure explains a phenomenon relative to other measures, should be documented to judge the utility of the new construct and the measure (Haynes & Lench, 2003). Therefore, the construct of learning potential should be validated within a network of supposedly related measures, and DA should be examined for its incremental predictive validity.

Despite its theoretical appeal and potential utility, DA research poses two challenges. First, because the term *dynamic assessment* is used to encompass all different approaches that incorporate instruction into assessment to measure learning potential, one of the challenges is that there is no single definition of what DA measures (Caffrey, Fuchs, & Fuchs, 2008; Grigorenko, 2008; Jitendra & Kameenui, 1993). Similarly, there is lack of a clear distinction in the literature between the definition of intelligence and learning potential, despite their different theoretical orientations. Unfortunately, this vagueness in learning potential constructs could result in the jangle fallacy (Kelly, 1927). The jangle fallacy refers to a situation when different terms are used to explain the same or similar phenomenon. The jangle fallacy can be problematic because it obstructs the synthesis of research findings. In addition, given the fact that it takes resources and effort to develop and validate a measure, committing the jangle fallacy may negate the effort put forth. One may be developing a measure (e.g., DA) that assesses the same construct that already exists (e.g., intelligence).

The definition of learning potential shows a considerable overlap with the definition of intelligence in the literature. Although an exact definition of what DA measures varies across different DA approaches, learning potential is commonly defined either as the capacity to master and reapply knowledge and skills taught during instruction (Grigorenko & Sternberg, 1998), or

as the responsiveness of the learner to instruction (Lidz, 1987). Proponents of DA assert that learning potential is distinct from intelligence. However, intelligence is often and commonly defined as the capacity to learn from experience (Sternberg & Detterman, 1986). If we substitute experience with instruction, the definition looks nearly identical to that of learning potential. The overlap in the definition can also be seen with the graduated prompts approach. The key concept of the graduated prompts approach is learning and transfer, which is an ability to use learned information in a variety of contexts, and the outcome is regarded as the efficiency of learning (Campione & Brown, 1987; Campione, Brown, & Bryant, 1985; Day, Engelhardt, Maxwell, & Bolig, 1997). Likewise, intelligence can be understood as how one learns from experience and how quickly one learns (Gottfredson, 1997). It is possible, then, that the same construct, ability to learn, is being labeled with two different terms (i.e., learning potential and intelligence). Thus, it is important to determine whether DA is a valid measure of learning potential rather than just another measure of intelligence. These concerns about whether intelligence and learning potential are qualitatively different have also been raised by Grigrenko and Sternberg (1998) stated as follows:

> Another concern that needs to be investigated is whether the way learning potential is defined in this paradigm (graduated prompts) is really qualitatively different from the traditional way ability is defined in the static-testing paradigm, or whether it is another measure of it that complements (but is nonidentical to) an existing measure. (p. 95)

Second, in addition to the challenge of validating learning potential as a distinct psychological entity, another challenge is that DA needs to show incremental validity to demonstrate its utility as a psychological construct. It has been noted that some innovative

constructs that make intuitive sense have frequently failed to add value when compared with preexisting measures (e.g., Sanders, Lubinski, & Benbow, 1995). Verifying the incremental validity of DA is important for two reasons. First, DA should be tested against other cognitive indicators that could affect performance on DA. Typically, a single measure taps many cognitive processes to varying degrees. For example, performance on math word problems may be affected by students' attention, reading, as well as their math skills. Thus, controlling for the sources of construct-irrelevant variance and showing incremental validity of the test is an important step to evaluate the validity of test score interpretation. Second, verifying the incremental validity of DA by showing its value above and beyond the competing predictors would establish the degree to which it might be practically useful in educational settings. DA has been criticized for being labor intensive in administration (Jitendra & Kameenui, 1993). If DA does not add a significant amount of information to the existing measures in explaining students' academic performance, the benefits of DA may not outweigh its costs.

Prior Dynamic Assessment Studies

To contextualize the present study, two lines of DA research that utilize the graduated prompts approach are reviewed. First, prior research examining the construct validity of DA in relation to intelligence or other constructs measured by static assessments (e.g., comparing static computation vs. dynamic computation) using factor analytic methods is reviewed. Second, results from previous studies that have explored the incremental validity of DA for explaining basic word reading skills or arithmetic outcomes are summarized.

Empirical evidence of the construct validity

Three studies were identified that used factor analytic methods to validate DA as a distinct construct from constructs measured with static assessments, including intelligence. First, Swanson and Howard (2005) measured information processing potential with two sets of DA, both employing verbal working memory tasks. One DA measured phonological working memory (rhyming task) and the other measured semantic working memory (digit/sentence). In both DA tasks, students were asked to retrieve information after performing an interference task. When students failed to accurately retrieve the information, students were given a series of graduated prompts based on the forgotten information. These working memory tests were given in three different conditions: initial, dynamic, and maintenance. Four scores were derived, three of which were the highest scores obtained during initial (initial score), dynamic (gain score), and maintenance condition (maintenance score), respectively. The fourth score was the probe score that was the total number of hints provided by the examiner during the dynamic testing trials.

The authors tested a five-factor model using a common factor analysis. These five factors were: static semantic, static phonological, verbal IQ, semantic responsiveness (i.e., semantic learning potential), and phonological responsiveness (i.e., phonological learning potential). Extracting learning potential factors separately for phonological and semantic tasks using probe and gain scores yielded a good fit to the data. These learning potential factors were found to be distinct from static semantic and phonological working memory, as well as verbal IQ. Results indicate that, within a task domain, empirical data were well represented by the model with a learning potential factor that was distinct from other factors comprising static measures. However, one weakness of this study was that nonverbal IQ, which could affect performance on

DA, was not part of the analysis. Also, their model was not compared with any competing models that may have been more parsimonious.

Fuchs et al. (2008) extended Swanson and Howard's study in three ways. First, based on prior findings that the more a DA task aligns with the academic content, the higher its predictive power, the authors developed DA of algebraic learning instead of basic cognitive skills. Second, Fuchs and colleagues included a variety of measures that could potentially influence the performance on DA including nonverbal IQ and static measures external to the DA tasks. Finally, the authors used structural equation modeling to test several competing measurement models. The DA of algebraic learning contained three algebraic skills, each of which had five levels of graduated prompts that were provided until mastery was shown. Three indicators of learning potential in algebra were used to create a latent factor (number of prompts for each skill). Researchers then compared models to see whether DA was a separate factor from language (including verbal IQ), attention, calculation, nonverbal reasoning (i.e., nonverbal IQ), and word problems. Results showed that a six-factor model was a better fit than the five-factor models (DA collapsed with one of other factors) and the one-factor model (general factor). Similar to results found by Swanson and Howard, Fuchs and colleagues found that learning potential for algebra was distinct from static measures of math (i.e., calculation, word problem solving skills) and from domain general abilities such as attention, nonverbal reasoning, and language.

Despite some consistency in findings in the above two studies, another study by D. Fuchs et al. (2011) produced somewhat different results. In this study, DA was developed for three decoding skills with graduated prompts embedded within each skill. Using the sum of the total number of prompts given across the three decoding skills, they conducted exploratory factor analysis with other language and reading related variables as well as measures of attention and

behavior. Results supported a three-factor structure in which DA loaded onto the first factor along with language, non-speeded reading, and IQ; the other two factors were behavior and rapid naming. The discrepancy between these results and previous ones might be due to several factors including the domain of the DA tasks (working memory and math vs. reading), the methods used to index learning potential (latent factor of DA vs. manifest variable), and the analytic method (confirmatory vs. exploratory). A limitation of this exploratory approach is that it explained what DA has in common with other measures, but did not provide information about how learning potential for decoding is different from static decoding skills or intelligence.

Overall, given the small number of studies that differ in various dimensions, limited empirical evidence exists to suggest that DA measures learning potential as a construct that is distinguishable from what can be measured by static assessments, including intelligence. Therefore, further studies (including the present one) are needed to examine the construct validity of DA.

Empirical evidence of the incremental validity

Evaluating the incremental validity of a measure depends on which criterion measures are being explained (Haynes & Lench, 2003). Accordingly, in the literature review of DA, Caffrey et al. (2008) reports that predictive validity of DA differs across types of criterion measures. Therefore, criterion for prediction was limited to word reading and arithmetic skills because they are critical milestones of reading and math development in earlier grades and because DA has been suggested as useful for supplemental screening or as an early identification tool for academic difficulties.

DA as a predictor of word reading. In the same study previously described, Swanson and Howard (2005) concurrently predicted word reading with five factors based on their factor

analysis results. Results from hierarchical regression indicated the phonological responsiveness factor did not significantly add to the prediction of word reading above the other factors. However, the semantic responsiveness factor was a significant predictor of word reading above the other factors. As a whole, the five factors accounted for 44% of the variance in word reading. When the two responsiveness factors were removed, the variance explained reduced to 38%. Thus, approximately 6% of the variance in word reading was accounted for by the responsiveness factors. However, given that the predictive ability of DA is higher when using domain specific tasks, this percentage may be higher when reading is assessed.

Several studies focusing on phonological awareness (PA) and decoding tasks in DA were identified in the reading domain. Spector (1992) investigated how well DA of PA predicts later word reading beyond three static measures of PA (i.e., phoneme segmentation, deletion, and invented spelling). DA in this study was adapted from a commonly used static PA measure that uses a form of graduated prompts with six levels. Responsiveness was operationalized as the total number of prompts provided. Results of a multiple regression provided strong support for the predictive validity of DA beyond all three static PA measures in word reading after controlling for verbal ability. In fact, DA was the only significant predictor of later word reading, and it explained 21% of variance in word reading. Thus, Spector's DA demonstrated incremental validity over static PA measures in word reading.

Another study that used PA tasks in DA was conducted by Bridges and Catts (2011). They used a sequence of three graduated prompts: first repeating the item, then emphasizing the sounds, and finally using pictures. They compared the predictive validity of their DA, first to its static version (score without prompts), and then to the common screening tool used in schools. Results from hierarchical regression indicated that DA only explained 4% more for word reading

and 9% more for decoding after controlling for the static version of PA task. Similarly, when compared to the widely used screening tool, DA accounted for a nonsignificant amount of the variance in word reading (2%) and 5% of the unique variance in decoding. Overall, less than 5% of the total variance explained was uniquely attributable to DA. Although Spector's (1992) study demonstrated promising results of DA using domain-specific tasks, overall findings are consistent in terms of the small amount of variance uniquely explained by DA in word reading skills. Additionally, all of these studies compared DA to only a small set of measures. The unique but small variance explained by DA may not be present when compared to a wide range of word reading predictors.

Some studies have specifically investigated the predictive validity of DA that focuses on decoding, which is a more proximal indicator of word reading than PA, for explaining later word reading outcomes. These studies used a stringent criterion for establishing the incremental validity of DA by comparing DA to a number of predictors that could potentially influence word reading outcome. Fuchs et al. (2011) also investigated the predictive validity of DA of decoding for predicting future word reading. The authors found that decoding DA was a significant predictor of word reading, even when controlling for the static measures of decoding. Approximately 4% of the variance was uniquely explained by DA depending on whether the outcome was a timed or untimed measure. Even when the other predictors were added to the model, including rapid naming, PA, language, behaviors, and IQ, DA of decoding still explained 2.3% of variance in untimed word reading, but DA lost predictive power for timed word reading.

An extension study of Fuchs et al. (2011) was conducted by Cho, Compton, Fuchs, Fuchs, and Bouton (in press). The authors used the same DA as Fuchs et al. to predict responsiveness to supplemental small group tutoring. Criteria for prediction were the final level and growth of

word reading fluency during small group phonics based intervention. DA was competed with three sets of predictors: both timed and untimed static decoding measures, final level and growth of word reading fluency during general classroom instruction, and four precursors of reading including PA, rapid letter naming, oral vocabulary, and IQ. Results from individual growth modeling indicated that DA was a significant predictor of final level and growth, uniquely explaining 3% to 13% of the variance in responsiveness to small group tutoring depending on the competing predictors.

As a whole, the present review suggests that DA measures have incremental validity in explaining or predicting word reading and its growth (Bridges & Catts, 2011; Cho et al., in press; Fuchs et al., 2011; Spector, 1992; Swanson & Howard, 2005). Although the consistent findings across these studies are that DA was a significant predictor of word reading skills, the amount of variance uniquely explained by DA was small. Additionally, the majority of these studies have mainly focused on comparing DA to its static version or to a small set of competing predictors. Such studies may overstate the utility of DA because they did not compare DA against other established predictors of word reading.

DA as a predictor of arithmetic. Only two studies were identified that specifically focused on arithmetic outcomes. Swanson and Howard (2005) concurrently predicted arithmetic using the same models applied to word reading outcomes. Similarly, only the semantic responsiveness factor was a significant predictor of arithmetic when controlling for the other factors, but not the phonological responsiveness factor. All five factors accounted for 54% of the variance in arithmetic and approximately 25% of the arithmetic variance was accounted for by the two responsiveness factors. This seemingly large amount of unique variance explained by

DA should be interpreted with caution because other predictors of arithmetic such as number sense and attention have not been taken into account.

Seethaler, Fuchs, Fuchs, and Compton (2011) developed the DA of balancing equations to measure students' learning potential for solving missing variables in nonstandard addition and subtraction expressions. Four types of equations were used and four or five graduated prompts were embedded for each type of DA. The authors predicted later computation performance using the sum of the total prompts provided. Unlike Swanson and Howard (2005), Seethaler and colleagues used a more stringent criterion for testing the incremental validity of DA. As competing predictors, two types of numerical competency measures (timed and untimed) were used in addition to domain-general cognitive functions such as verbal ability and nonverbal reasoning. Regression results indicated that DA was a significant predictor of later calculations beyond the competing predictors. As expected, commonality analysis results suggested smaller variance was uniquely explained by DA (2.8%) than that of Swanson and Howard.

Across word reading and arithmetic outcomes, results are promising. The reviewed studies documents unique, although small, predictive validity for DA beyond traditionally used static measures. However, less is known about the benefits DA when compared to wide range of other known predictors. Also, studies have rarely investigated whether DA in certain domain has incremental validity across different domains.

Present Study

Study purpose

This study comprises two major components. First, a DA of decoding that taps the basic cognitive processes involved in decoding (i.e., letter-sound correspondence learning, blending,

and rule-based decoding) was developed and its construct validity was tested. The purpose of this examination was to explore whether learning potential for acquiring skills necessary for decoding is a construct distinct from domain-general learning aptitude (intelligence) and from current decoding skills.

Second, incremental validity of DA of decoding predicting concurrent word reading skills was examined. Compared to prior studies, a more stringent criterion was used to test the incremental validity of DA. Not only was a comprehensive lists of competing predictors of word reading included, but also domain-general cognitive predictors, which are assumed to play key roles in the initial stages of learning novel skills. These various domain-general covariates were included to compete with DA. In addition, the differential predictive validity of DA for explaining real word reading (word recognition) and nonword reading (decoding) was examined.

Similarly, whether DA of decoding has incremental validity across domains was explored by examining whether it has additive value in explaining arithmetic performance. Previous work indicates that as DA tasks move closer to the relevant academic content, the utility of DA for predicting academic learning improves. Thus, except for Swanson and Howard (2005), the rest of the reviewed studies focused on predictive validity within a certain domain. However, we do not know whether learning potential is governed by a domain-general mechanism or whether learning potential in one domain can be generalized to other domains (Grigorenko, 2009). Unfortunately, prior work has rarely examined whether responsiveness in a certain domain predicts academic performance across domains.

Research questions and hypotheses

Research question 1. Is DA measuring learning potential for acquiring decoding skills distinct from conventional static measures of intelligence and decoding? DA is hypothesized to

be distinct from intelligence as well as from static decoding skills. And although distinct, these constructs are hypothesized to be moderately correlated.

Research question 2. Does DA of decoding have incremental validity in explaining word reading skills beyond what can be explained by the known predictors of reading and domain-general learning indicators? Precursors of word reading skills are well established, and they include: PA (for a review, see Bus & Ijzendoorn, 1999), and rapid automatized naming (RAN; Blachman, 1984; Compton, 2000; Wolf, 1986) including rapid letter naming (RLN) and rapid digit naming (RDN). The domain-general indicators of learning included in the present study were verbal intelligence, nonverbal intelligence, and attentive behavior. Based on the prior findings and the theory that DA adds unique information to what static measures are unable to provide, DA of decoding is hypothesized to provide unique additional information above and beyond these reading and domain-general predictors for explaining word reading skills, both word recognition and decoding. Further, DA is hypothesized to be a stronger predictor of decoding than word recognition.

Research question 3. Does DA of decoding have incremental validity across domains? As previously mentioned, the definition and nature of learning potential are vague. Grigorenko (2009) has also pointed out that whether learning potential is generalizable or domain specific is a contentious issue that has rarely been investigated. Some researchers have conceptualized learning potential measured via DA as *cognitive modifiability* that is governed by a domain-general mechanism (e.g., Campion & Brown, 1990). If learning potential is a domain-general learning mechanism, DA in one domain may predict performance in other domains. The present study included arithmetic as a cross-domain criterion while taking into account predictors of math such as number sense. Learning potential is hypothesized to be domain-specific for the

following two reasons. First, Day et al.'s (1997) study suggests the domain specificity of learning potential. Comparing DA of similarities (i.e., verbal domain) and block design (i.e., nonverbal domain) tasks, the researchers found that responsiveness indices of the two tasks were not correlated. The second reason for hypothesizing domain-specificity of DA is that DA has shown differential patterns of predictive validity for explaining two different outcomes even within the same domain (e.g., Bridges & Catts, 2011; Fuchs et al., 2011; Seethaler et al., 2011).

CHAPTER III

METHOD

Participants

One hundred and twelve first grade students with various levels of reading ability participated in this study. Based on teacher reports, only those students whose native language is English were included. Demographic information of the participants is summarized in Table 1.

Table 1.
Demographics of the Participants (N=112)

	Frequencies (%)
Gender	
Male	64 (57.14)
Female	48 (42.86)
Race	
African American	43 (38.74)
Caucasian	55 (49.55)
Others	13 (11.61)
Free/Reduced Lunch	
No	53 (47.32)
Yes	59 (52.68)
IEP	
No	102 (91.07)
Yes	10 (8.93)
	M (SD)
Age	6.718 (.319)

Measures

Dynamic assessment of decoding

Three essential skills required for decoding development were assessed in the DA: learning symbol-sound correspondence, blending sounds, and inferring decoding rule. The learning and mastery of each skill was assessed using the following tasks. For the symbol-sound

correspondence task (DA 1), students were asked to make connections between six new symbols (adopted from Chinese characters) and their corresponding English sounds (s, m, t, p, f, a). For the blending task (DA 2), students were asked to read consonant-vowel-consonant (CVC) words written with the new symbols they learned in DA 1. For the inferring decoding rule task (DA3), students were required to discover the 'silent e' rule and read CVCe words.

For each task, students were given multiple trials to master the skill. Each learning trial was composed of two parts: Instructional prompts and a six-item posttests. The general procedure was as follow: 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 the provision of instructional prompts to help them master the skill. If students failed again, they were given the next learning trail with more explicit prompts. Increasingly explicit prompts were given until the student 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 unadministered items. If students did not show evidence of learning even after all the levels were provided, the tester stopped administration. In this case the students received a 0 score for the remaining unadministered tasks. Students were provided 9 learning trials at most for DA 1, 4 trials for DA 2, and 5 trials for DA 3.

Description of the increasingly explicit prompts for each DA task was as follows: In DA 1, students were presented with the novel symbols and asked to say the sounds. Initially for the first 5 trials, students were provided only with corrective feedback; then, students were provided with a key word representing its sound (e.g., /a/ as in apple); then, provided with partial picture clues (e.g., apple-like picture that resembles the /a/ symbol); then, provided with complete picture clues (e.g., apple picture); and finally, students were asked to trace over the symbol with

their finger. In DA 2, students were presented with CVC words (i.e., sam, fat) written with the novel symbols. First, the tester read the word; second, the tester modeled sounding out (blending); third, the tester modeled tapping out by breaking down each individual sounds, and then sounded out; and finally, along with the sounding and tapping out, 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 words and CVCe words (i.e., sam/same, fat/fate). To represent CVCe words, another novel symbol was added that represents 'e' to CVC words. First, the tester read the word; second, the tester tapped out the sound in the word; third, the tester directed students' attention to the middle sound and instructed them that the middle sound of the CVCe word was different from 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.

After each of the instructional prompts, students' learning was assessed using the posttest. The test comprised six items. The items were repeated across the tests but were presented in a random order in each mastery test. Mastery test items were not used for instructional prompts. Outcome was the total number of correct items across the mastery tests. Internal consistency was .71 across all mastery tests and .50 for DA 1, .93 for DA2, and .95 for DA 3. Somewhat low internal consistency for DA 1 mastery suggests that paired associate learning may be dependent on the specific feature of the symbol.

The DA of decoding used in this study was modified from the measure developed by D. Fuchs and colleagues (2007), and incorporated a paired associative learning (PAL) task influenced by Elbro, Daugaard, and Gellert's DA design (2012). I modified the Fuchs et al.'s DA of decoding because it has been shown to have incremental validity for predicting later word

reading outcomes (e.g., Cho et al., in press; Fuchs et al., 2011) and shown to improve classification accuracy (Compton et al., 2010) when added into a comprehensive set of reading predictors of (for a detailed description of the previous DA, see Fuchs et al., 2007).

Based on the fact that Fuchs et al.'s (2011) DA showed floor effects and had shared variances with reading and language skills measured with static assessments, modifications were made in four ways. First, I centered the DA tasks on the three basic cognitive skills required for decoding: (a) learning novel symbol-sound correspondences, (b) blending sounds, and (c) learning rule-based decoding. I expected this would make the test more sensitive for discriminating individual in the lower end of the distribution because it included prerequisite skills required for decoding (letter-sound knowledge). Second, I included five trials of PAL in DA1. One reason for inclusion was that PAL has been considered a type of DA because it provides corrective feedback (e.g., Elbro et al., 2012). Another reason for the inclusion was that there have been reports indicating PAL is a significant predictor of word reading (e.g., Warmington & Hulme, 2012). Third, I used novel symbols instead of alphabet letters so that it is less dependent on prior knowledge about letter-sound correspondences and on existing decoding skills. As suggested by Sternberg and Grigorenko, the task should be novel in order to truly assess how the child may learn and transfer learning to a novel setting (1999). Thus, I sought to make the DA less dependent on prior reading skills and more dependent on learning potential by using novel symbols. Fourth, I attempted to reduce the language load in the DA tasks by minimizing verbal explanation. Most of the prompts in the current DA use visual cues and modeling.

Reading measures

A comprehensive set of static early indicators of reading were collected to include as predictors of reading skill: RLN, RDN, and PA. Two measures of decoding (timed and untimed) and word recognition (timed and untimed) were assessed as reading outcome measures.

Rapid automatized naming. The Comprehensive Test of Phonological Processing: Rapid Letter Naming (CTOPP: RLN; Wagner, Torgesen, & Rashotte, 1999) and Rapid Digit Naming (RDN) were used to measure the speed at which students can name two sets of 36 letters/numbers. According to the manual, the test-retest reliability for ages 5-7 is .97.

Decoding. The Woodcock Reading Mastery Test-Revised/NU: Word Attack (WRMT-R/NU: WAT; Woodcock, 1998) was used to measure untimed pseudoword reading in isolation. The manual reports the split-half reliability for first grade students as .94. In addition, the Test of Word Reading Efficiency: Phonemic Decoding Efficiency (TOWRE: PDE, Torgesen, Wagner, & Rashotte, 1997) was used to measure decoding accuracy and fluency. Test-retest reliability reported in the manual is .86 for the first grade sample.

Word recognition. The WRMT-R/NU: Word Identification (WID; Woodcock, 1998) was used to measure untimed real word reading in isolation. The split-half reliability from the manual is .98 for first grade students. Also, the TOWRE: Sight Word Efficiency (SWE; Torgesen, Wagner, & Rashotte, 1999) was used to measure students' word recognition fluency. Test-retest reliability reported in the manual is .93 for the first grade sample.

DA word list. Students were asked to read a list of 12 words (6 CVC and 6 CVCe) that were used in the DA mastery tests. This measure was included because it provided a baseline for static decoding skill that may affect the performance on DA mastery test. Internal consistency was .89 for the current sample.

Math measures

Two measures assessing number sense were included in addition to the DA of decoding for predicting arithmetic. Also, one standardized measure of arithmetic and two curriculum-based measures of computation (timed and untimed) were used as outcome.

Number sense. Quantity Discrimination (QD; Chard et al., 2005; Research Institute on Progress Monitoring, 2009) was used to measure students number sense. Test-retest reliability reported from prior research is .85 -.99 (Clarke, Baker, Smolkowski, & Chard, 2008). In addition, Number Sets (NS; Geary, Bailey, & Hoard, 2009) was used to assess the fluency with which students can identify and process quantities represented by Arabic numerals and object sets. Students performed four trials. Each trial yielded hit, miss, correct rejection, and false positives. Scores subtracting false alarm from hit scores for each form were used in the analysis.

Arithmetic. The arithmeticsubtest of the Wide Range Achievement Test-3 (WRAT-3: Wilkinson, 1993) was used. This consisted of 15 oral questions and 40 written arithmetic items. According to the manual, internal consistency reliability is above .81 for age 6 -7. In addition, the first grade math curriculum-based measure (CBM; Fuchs, Hamlett, & Fuchs, 1990) was used to assess students' addition and subtraction accuracy and fluency. Alternate form reliability reported in the manual exceeds .89.

Domain-general measures

Intelligence, attentive behavior, and phonological awareness are considered domain-general predictors of academic outcome. PA has long been considered to predict reading development. However, PA has recently gained attention as important predictor of arithmetic, especially for simple problems where math facts are retrieved from memory (De Smedt, Taylor, Archinald, & Anasari, 2010; Fuchs, Compton, Fuchs, Paulsen, Bryant, & Hamlet, 2005; Hecht,

Torgesen, Wagner & Rashotte, 2001). Thus, it is considered as domain-general predictor of learning in this study.

Intelligence. Wechsler Abbreviated Scale of Intelligence: Matrix Reasoning (WASI: MR; Wechsler, 1999) was used to measure nonverbal intelligence. The split-half reliability reported in the manual is over .85 for students age 6-7. According to the manual, the intercorrelation of MR with performance IQ is .87 and full scale IQ is .86. WASI-Vocabulary (Wechsler, 1999) was used to measure verbal intelligence, which includes expressive vocabulary, verbal knowledge, and foundation of information. The split-half reliability reported in the manual is over .86 for students ages 6-7. According to the manual, the intercorrelation of Vocabulary with verbal IQ is .93 and full scale IQ is .87.

Attentive behavior. In addition, the Strengths and Weaknesses of ADHD-Symptoms and Normal-Behavior (SWAN; Swanson et al., 2004) scale was used to assess students' attentive behavior. Only the items 1-9, which relates to the inattention were used. Internal consistency reliability coefficient (alpha) for the present sample is .97.

Phonological awareness. The CTOPP: Elision (Wagner et al., 1999) was used to assess student's phonological awareness. Children were asked to say a word, then to say the word after deleting a specified part of the word. According to the manual, the test-retest reliability for ages 5-7 is .88.

Procedures

In October and November of first grade, all of the measures were administered to students during two one-hour testing sessions. Teachers completed demographic form (including date of birth, ethnicity, free/reduced lunch status as a proxy for SES, placement information,

ELL status, and native language) and SWAN (Swanson et al., 2004). 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; 20% of tapes were selected randomly, stratifying by tester, for procedural fidelity and scoring reliability by an independent scorer. Procedural fidelity was above 97% across all assessment. For scoring reliability, measure of reading and oral vocabulary exceeded 90% reliability.

CHAPTER IV

DATA ANALYSES AND RESULTS

A series of confirmatory factor analyses were carried out to assess how well DA of decoding represents the target construct of learning potential distinct from general intelligence and from statically measured decoding ability (Research Question 1). Then, a series of structural equation models were used to examine the relationships (a) between a latent variables representing learning potential, decoding, and word recognition abilities (Research Question 2) and (b) between learning potential and arithmetic performance (Research Question 3) controlling for a wide range of competing predictors. These analytic procedures were selected because they represent a theory-driven approach for the evaluation of a priori hypotheses about causal relations among the variables. Raw scores were used in the analysis and data were analyzed using the Mplus 6.0 (Muthén & Muthén, 1998-2010). Judgment of statistical significance was based on the unstandardized output, but standardized coefficients are reported to aid the interpretation of the results.

Descriptive Statistics and Preliminary Analyses

Descriptive statistics (means, standard deviations) and zero-order correlation are reported in Table 2. No outliers (e.g, > 3.5 SDs from the mean) were identified. Because maximum likelihood (ML) estimation assumes univariate and multivariate normality, violation of these assumptions can lead to underestimation of standard errors and overestimation of the chi-square values. Therefore, distributional estimates of skew and kurtosis were tested using STATA 11.0

(StataCorp, 2009). Several variables showed significant skew and kurtosis. Mardia's normalized coefficient of multivariate normality was also checked to ensure the data met the assumptions of ML. Mardia's statistic for multivariate normality revealed significant multivariate nonnormality of skewness, 86.21, χ^2 (1330) = 1656.68, but not for kurtosis, 403.02, χ^2 (1) = .57. Thus, models were run with MLM developed by Satorra and Bentler, which uses robust standard errors and mean-adjusted chi-square test statistics that are robust to non-normality (1988). Satorra-Bentler chi-square incorporates a scaling correction factor for the chi-square statistic when distributional assumptions are violated and it has been shown to be the most reliable method for evaluating mean and covariance structure models (Curran, West, & Finch, 1996).

Table 2.

Descriptive Statistics and Zero-order Correlation among Manifest Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1. DA1	-																						
2. DA2	.37	-																					
3. DA3	.35	.53	-																				
4. MR	.30	.30	.40	-																			
5. VOC	.23	.40	.50	.49	-																		
6. ATT	.09	.29	.33	.30	.48	-																	
7. PA	.27	.42	.52	.35	.50	.17	-																
8. RDN	-.13	-.32	-.27	-.11	-.29	-.41	-.09	-															
9. RLN	-.12	-.34	-.30	-.20	-.31	-.44	-.21	.85	-														
10. QD	.06	.32	.31	.18	.44	.51	.30	-.63	-.64	-													
11. NS5a	.11	.35	.28	.35	.55	.49	.25	-.33	-.27	.56	-												
12. NS5b	.22	.47	.46	.41	.64	.48	.44	-.33	-.32	.61	.71	-											
13. NS9a	.05	.15	.24	.11	.22	.38	.12	-.33	-.19	.40	.55	.39	-										
14. NS9b	.01	.30	.34	.33	.50	.50	.36	-.35	-.31	.52	.66	.69	.63	-									
15. DAW	.15	.42	.62	.38	.67	.47	.61	-.33	-.42	.43	.38	.56	.11	.52	-								
16. WAT	.24	.50	.64	.42	.57	.47	.66	-.28	-.37	.37	.36	.52	.22	.55	.80	-							
17. PDE	.24	.40	.51	.37	.51	.47	.53	-.36	-.40	.40	.35	.52	.27	.57	.70	.83	-						
18. WID	.23	.43	.54	.37	.56	.52	.59	-.39	-.48	.44	.35	.54	.22	.56	.79	.88	.84	-					
19. SWE	.14	.32	.41	.35	.53	.48	.47	-.44	-.50	.44	.35	.52	.29	.56	.71	.75	.84	.90	-				
20. AR_o	.19	.50	.44	.36	.55	.41	.46	-.48	-.49	.56	.46	.60	.23	.54	.59	.59	.48	.57	.50	-			
21. AR_w	.12	.29	.34	.33	.50	.41	.45	-.32	-.28	.45	.47	.55	.34	.60	.54	.62	.57	.59	.56	.51	-		
22. CBM2	.04	.30	.30	.16	.33	.36	.25	-.41	-.34	.54	.52	.64	.43	.64	.43	.40	.47	.44	.49	.52	.47	-	
23. CBM5	.03	.35	.47	.29	.56	.40	.42	-.29	-.29	.48	.57	.68	.43	.67	.64	.57	.51	.54	.52	.58	.59	.62	-
M	44.44	18.95	15.88	52.61	52.87	4.45	8.46	1.40	1.34	31.04	7.71	7.51	5.58	5.69	8.68	16.05	13.94	38.91	34.17	13.71	3.80	8.96	14.65
SD	7.86	7.54	11.60	11.33	12.03	1.45	4.61	.34	.34	8.05	2.80	3.82	2.50	3.83	3.41	8.67	8.72	14.83	14.62	1.49	1.36	4.92	6.03
Min	19	0	0	35	20	1	0	.53	.43	5	-1	-5	-1	-6	0	0	0	1	3	8	1	0	0
Max	54	24	30	77	80	7	19	2.4	2.32	50	13	16	10	15	12	38	45	73	67	15	9	24	25

Note. Coefficients greater than .18 is significant at α=.05; DA1-3it = total number of item correct for DA 1-3; DA1-3lv = total number of prompts given for DA 1-3; MR = matrix reasoning (nonverbal intelligence); VOC = vocabulary (verbal intelligence); ATT = attentive behavior; PA = phonological awareness; RDN = rapid digit naming; RLN = rapid letter naming; QD = quantity discrimination; NS5a = first probe of number set for 5; NS5b = second probe of number set 5; NS9a = first probe of number set for 9; NS9b = second probe of number set 9; DAW = DA word list; WAT = word attack; PDE = phonemic decoding efficiency; WID = word identification; SWE = sight word efficiency; AR_o = Wide Range Achievement Test 3rd Arithmetic oral section; AR_w= Wide Range Achievement Test 3rd Arithmetic written section; CBM2 = computation curriculum based measure timed for 2 min; CBM5 = computation curriculum based measure timed for 5 min.

Model Fit Indices

Model fit in these analyses were evaluated based on multiple fit indices including absolute fit indices and incremental fit indices. As recommended by Hu and Bentler (1998), standardized root-mean-square residual (SRMR; absolute fit) was reported with other supplementary fit indices such as comparative fit index (CFI; incremental fit Type 3) and nonnormed fit index (incremental fit Type 2) also known as Tucker-Lewis index (TLI). These fit indices are sensitive to model misspecification, less sensitive to distribution and sample sizes, and stable across different estimation methods in ML (Hu & Bentler, 1998). CFI is an especially good index with small sample size. TLI penalizes for model complexity and allows researchers to evaluate models for parsimony. When evaluating model fit, greater weights were given in the following order: SRMR, CFI, and TLI. A cutoff value of .08 is used for SRMR (Hu & Bentler, 1998; Browne & Cudeck, 1993). Although it is recommended that TLI and CFI be close to .95, models with values close to .90 were considered acceptable (Bentler, 1992; Kline, 2005). In addition, chi-square difference tests, adjusted with the scaling correction factor, were conducted to inform whether the base model provides better fit than competing models. The base model was considered to yield better fit because it includes a larger number of estimated parameters (i.e., less restrictive). However, if nested models (more restrictive) fail to have significantly poorer fit than the base model, the more parsimonious model was selected.

Research Question 1

A series of confirmatory factor analyses were conducted to address the issue of construct. The first latent factor, learning potential, composed of indicators from all three DA tasks (Figure 1): DA 1, DA 2, and DA 3. The second factor, general learning aptitude, composed of WASI

Vocabulary (verbal intelligence) and MR (nonverbal intelligence). The third factor, decoding, included WAT (untimed decoding), SWE (timed decoding), and DA word list. Reading DA word list was considered as a decoding process because it comprised words that follow the decoding rule (e.g., CVC, and silent *e* rule)

Four hypothetical models were tested. Model 1 had three separate, correlated factors (learning potential, decoding, general learning aptitude). This model assumed that potential for learning to decode is different from present level of decoding skills and from domain general learning aptitude. This model supposedly has the best fit because it has the fewest constraints. Thus, Model 1 served as a baseline from which to evaluate the fit of the other models.

Model 2 had two separate, correlated factors (learning potential + general learning aptitude, decoding) and assumed learning potential for decoding and general learning aptitude as a unitary construct. The rationale for this model comes from the concern of the Jangle fallacy that we might be labeling one construct (learning aptitude) with two labels, intelligence and learning potential. Model 3 had two separate, correlated factors (learning potential + decoding, general learning aptitude). According to this model, student's potential for learning to decode is not different from to how well one can decode. The last model, Model 4, assumed a single general factor. According to this model, there is a single underlying ability that is manifested in individual differences on learning to read in general.

Results showed that the three factor model (Model 1), with separate but correlated factors representing learning potential, general learning aptitude, and decoding, was a good representation of the data structure, $\chi^2(17) = 36.420$, $p = .004$, CFI = .960, TLI = .935, SRMR = .042 (see Model 1 in Table 3). Because alternate models were nested within the three factor model, adjusted chi-square difference tests were conducted. Adjusted chi-square values were

7.019 (Δ*df* = 2), 18.066 (Δ*df* = 2), and 22.92 (Δ*df* = 3) for Model 2 - 4, respectively. All of them were significant at a .05 alpha level; thus, Model 1 was preferred over the alternative models (Table 3). I therefore concluded that DA measures the construct of learning potential that is distinct from static decoding skills and general learning aptitude. Although separate, correlations of these latent constructs suggest that these three dimensions are highly correlated (Figure 1). All observed variable loaded reliably onto their respective factors in Model 1 (standardized coefficients of .420 - .959, *ps* < .001).

Table 3.
Fit Indices and Model Comparisons for the Competing Models for Research Question 1.

Model	*df*	χ^2	SCF	*p*	CFI	TLI	SRMR	$\Delta\chi^2$ Model 1
Three factor model								
Model 1: LP, IQ, DEC	17	36.420	.945	.004	.960	.935	.042	
Two factor model								
Model 2: LP + IQ, DEC	19	44.015	.969	.000	.949	.925	.050	7.019*
Model 3: LP + DEC, IQ	19	56.795	.964	.000	.923	.886	.063	18.066**
One Factor Model								
Model 4: LP+ IQ + DEC	20	60.801	.960	.000	.912	.877	.065	22.92**

Note. LP = learning potential; D = decoding; $\chi 2$ = Satorra-Bentler corrected chi-square statistics; SCF = scaling correction factor for Satorra-Bentler χ^2; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root-mean-square residual; $\Delta\chi^2$ = adjusted chi-square difference test
* $p < .05$
** $p < .001$

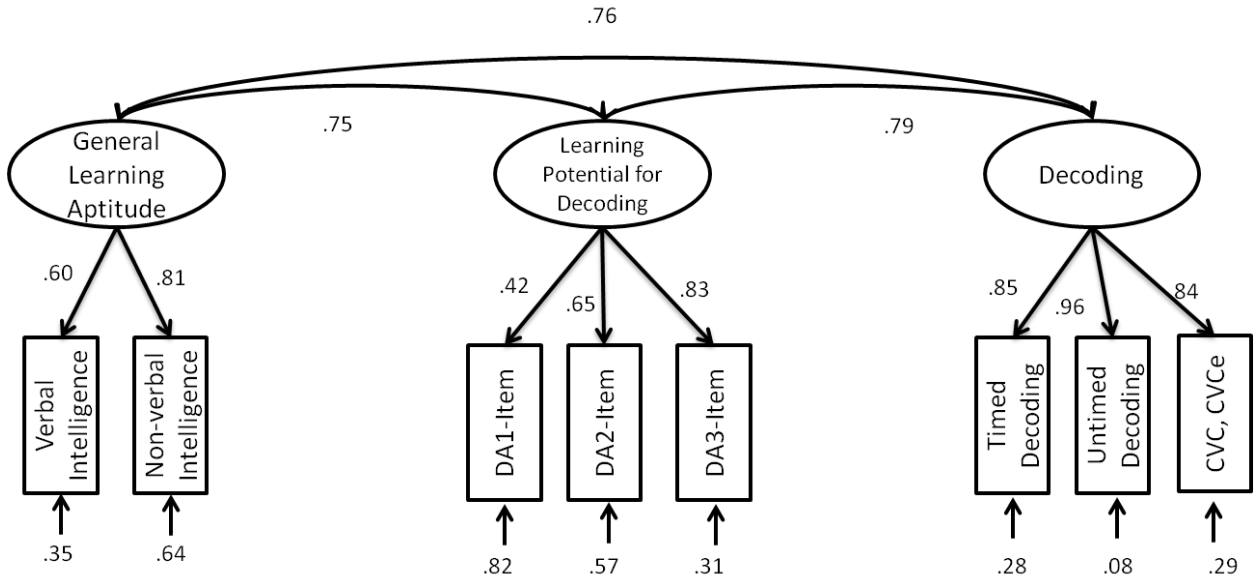


Figure 1. Three-factor model for research question 1. All coefficients are depicted in the figure are significant. Verbal intelligence (Wechsler Abbreviated Scale of Intelligence [WASI] Vocabulary [VOC]), nonverbal intelligence (WASI Matrix Reasoning [MR]), learning Potential (total number of correct items in mastery tests for DA1, DA2, DA3), untimed decoding (Woodcock Reading Mastery Test [WRMT] Word Attack, timed decoding (Test of Word Reading Efficiency [TOWRE] Phonemic Decoding Efficiency), cvc and cvce words (DA word list)

Research Questions 2 and 3

A series of structural equation modeling progressed in two stages. First, the measurement models were estimated separately for predictors and outcomes. Second, various relational models were tested using structural equation modeling. The model building process of the measurement models was guided by both theory and data. Initially, I assumed the most parsimonious predictor model with five latent factors including general learning aptitude (MR, VOC), attentive behavior (SWAN), learning potential (DA1, DA2, DA3), pre-reading (PA, RLN, RDN), and pre-math (PREM; QD, NS5a, NS5b, NS9a, NS9b). However, this measurement model did not fit the data structure well, $\chi^2(94) = 233.267$, $p = .000$, CFI = .884, TLI = .852, SRMR = .108. An alternative

model was estimated with three latent factors including learning potential, rapid automatized naming (RAN; RLN, RDN), and PREM as well as four manifest variables including nonverbal intelligence (MR), verbal intelligence (VOC), attentive behavior (SWAN), and phonological awareness (PA). Factors with only one indicator were entered in the model as manifest variables (Figure 2).

Theoretical justification for separating general intelligence into nonverbal and verbal domain was as follows: Nonverbal and verbal intelligence represent two distinct sub-dimensions in the hierarchical structure of the intelligence (e.g., Carroll, 1993; Cattell, 1971; Spearman, 1927). Nonverbal intelligence refers to the speed and accuracy of abstract reasoning for novel problems, whereas verbal intelligence refers to the accumulated knowledge and vocabulary from experience. These two separate dimensions of intelligence structure are empirically supported (Ryan et al., 2003; Weschler, 1999). In the case of the PA and RAN, although some have argued both skills fall under a unitary structure of phonological processing skills, others have argued for separable, though related, phonological ability constructs (e.g., Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993). PA is the awareness of and ability to manipulate the sound structure in spoken language, and RAN is the efficiency of the phonological code retrieval from a long-term storage. The speediness, cross-modality of stimulus and response, and retrieval of the phonological code make RAN distinguishable from PA. In fact, RAN and PA have been proposed as separate sources of deficits in reading development (e.g., Wolf & Bowers, 1999). This predictor measurement model provided an acceptable fit of the data, χ^2 (60) = 129.726, p = .000, CFI = .909, TLI = .861, SRMR = .063. All factor loadings for the latent factor were significant (standardized coefficient of .458-.920, *ps* < .001).

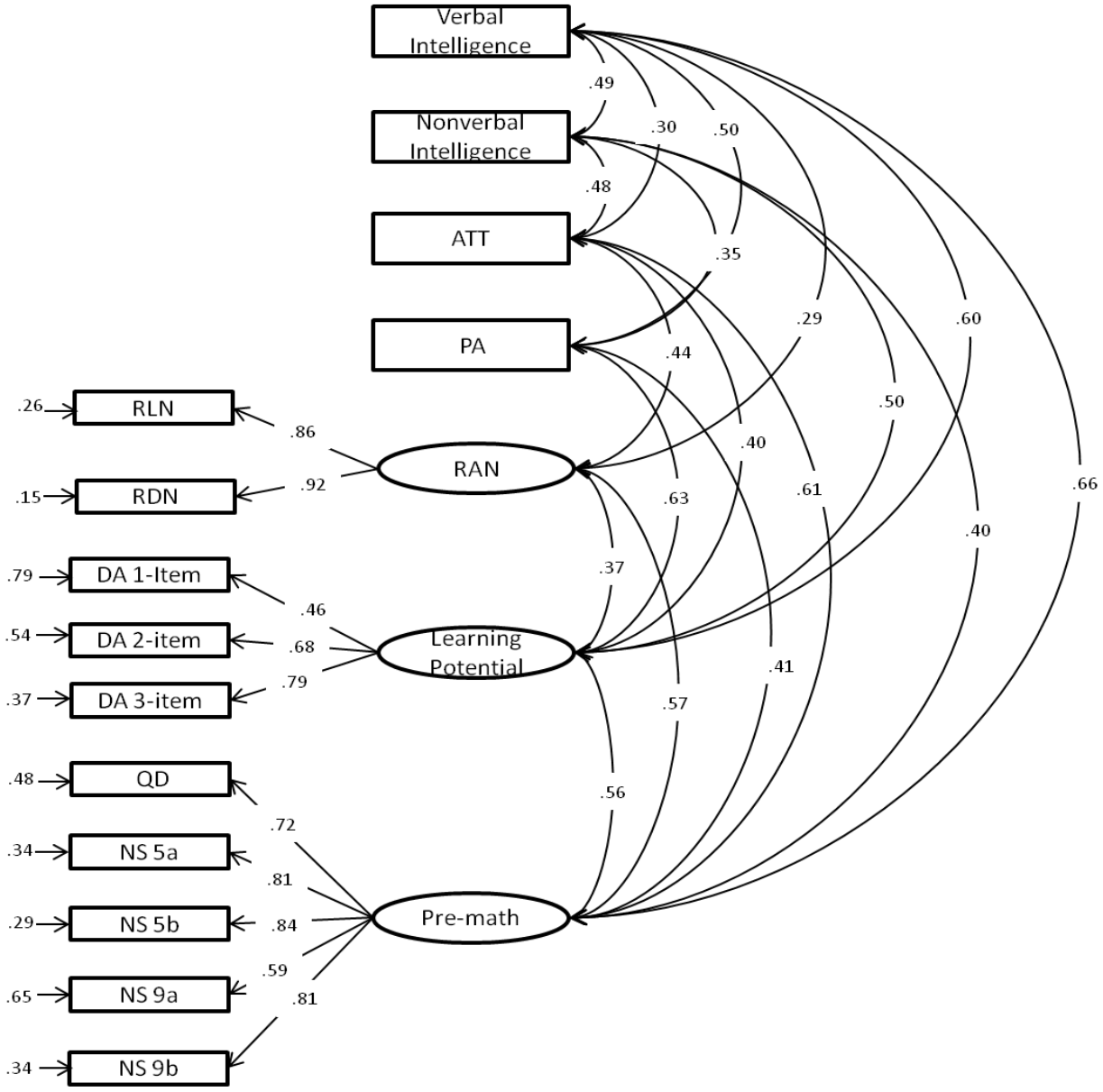


Figure 2. Measurement model for research questions 2 and 3 predictors. Only significant coefficients are depicted in the figure. The predictor model incorporated 7 dimensions of abilities: verbal intelligence (Wechsler Abbreviated Scale of Intelligence [WASI] Vocabulary [VOC]), nonverbal intelligence (WASI Matrix Reasoning [MR]), attentive behavior (SWAN Rating Scale), phonological awareness (Comprehensive Test of Phonological Processing [CTOPP] Elision), rapid automatized naming (CTOPP Rapid Letter Naming [RLN], Rapid Digit Naming [RDN]), Learning Potential (total number of correct items in mastery tests for DA1, DA2, DA3), predictors of math (Quantity Discrimination, Number Set [NS]5 set a, NS5 set b, NS9 set a, NS9 set b)

The measurement model for outcome included three correlated dimensions (Figure 3). The first factor, decoding, included WAT and PDE; the second factor, word recognition, included WID and SWE; and the third factor, arithmetic, included oral and written section of WRAT Arithemetic, 2-minute timed CBM (CBM 2), and 5-minute timed CBM (CBM 5). The three-factor solution represented the data structure well, χ^2 (11) = 55.697, p = .000, CFI = .951, TLI = .919, SRMR = .037. All observed variables loaded substantially and reliably onto their respective three-factors (standardized coefficients of .691-.922, *ps* < .001). An alternative two factor model comprised word reading (WAT, PDE, WID, SWE) and arithmetic (WRAT oral, WRAT written, CBM2, CBM5) also fit the data adequately. Yet, this alternative model fit significantly worse than the three factor model, adjusted $\Delta\ \chi^2$ (2) = 10.035, p < .05. Therefore, three-factor model was selected as a final outcome model.

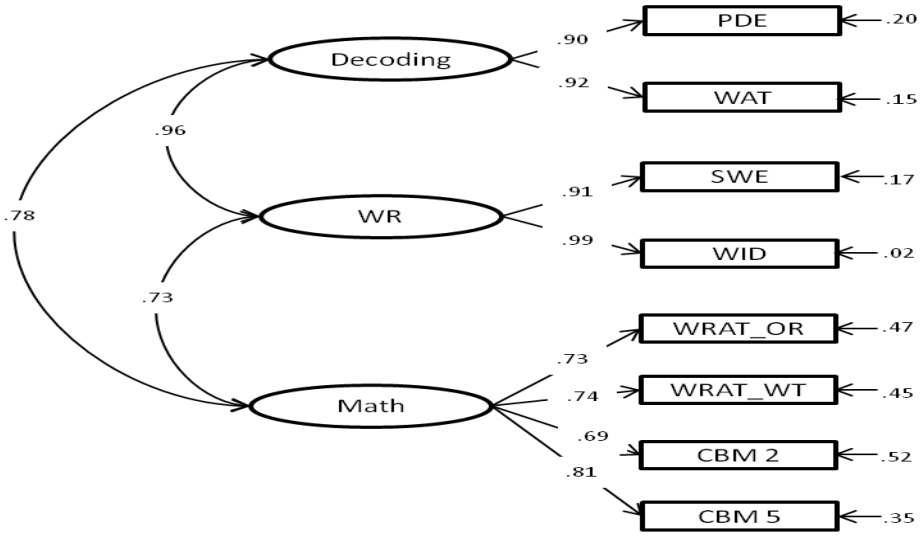


Figure 3. Measurement model for research questions 2 and 3 outcomes. All coefficients are depicted in the figure are significant. The outcome model incorporated 3 dimensions of academic skills: decoding (Woodcock Reading Mastery Test [WRMT] Word Attack, Test of Word Reading Efficiency [TOWRE] Phonemic Decoding Efficiency), word reading (WRMT Word Identification, TOWRE Sight Word Efficiency), arithmetic (Wide Range Achievement Test [WRAT] Arithmetic Oral Section, Written Section, 1st grade computation curriculum based measures [CBM] timed for 2 minutes and 5 minutes. WR = Word Recognition; Math = arithmetic.

The second stage of model building tested whether sequentially fixing the paths from learning potential to outcomes yielded significantly poorer fit than the comparison models. This procedure was used to examine whether learning potential is necessary in forecasting decoding, word recognition, and arithmetic performances. Four hypothetical models, depicted in Figure 4, were tested to represent the relationship between DA and the outcome measures.

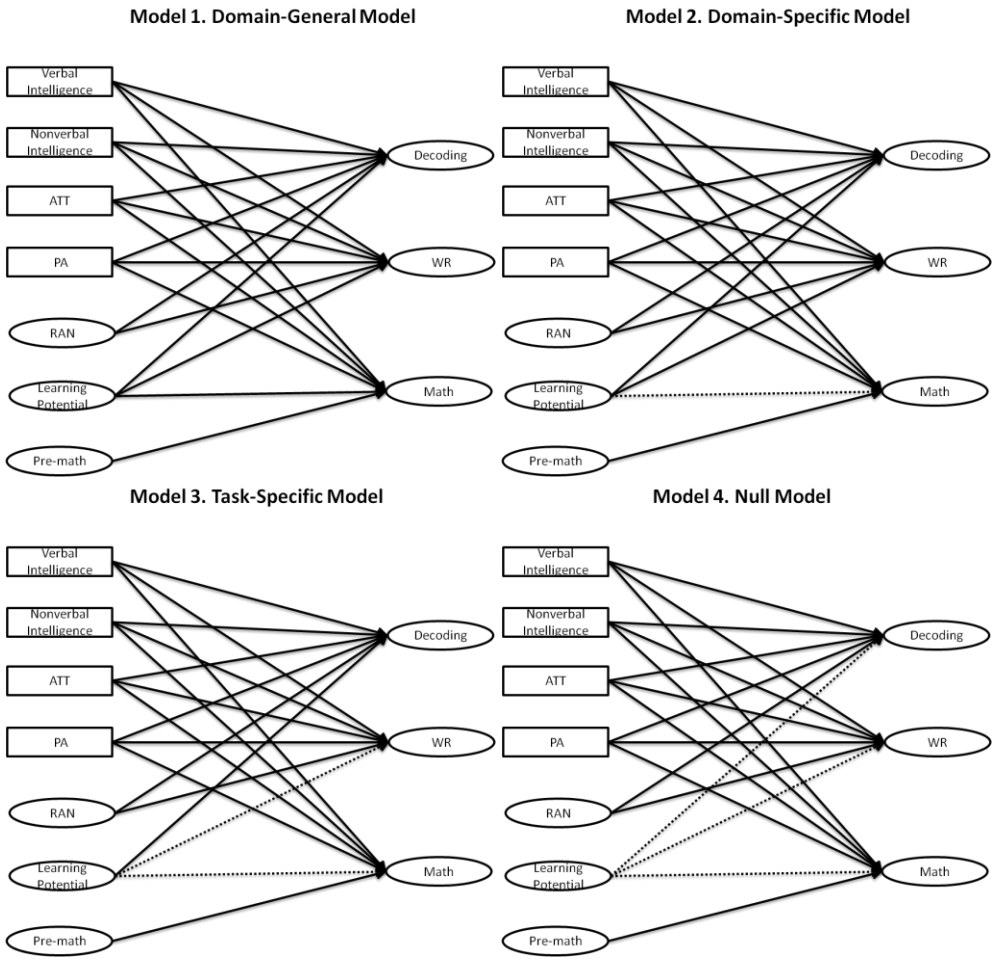


Figure 4. Four structural models explaining relations among predictor abilities and academic outcomes. Dotted line indicates the path coefficient is fixed to zero. The domain-general model is the least parsimonious structural, in which DA is included as a predictor for decoding, word recognition, and arithmetic outcome. Each of the subsequent (nested) models is more parsimonious. In domain-specific model, DA is included as a predictor only for decoding and word recognition. In task-specific model, DA is included as a predictor only for decoding. In the null model, DA is included as a predictor of none of the outcomes. WR = Word Recognition; Math = arithmetic

The base model (Model 1) was the least parsimonious structural model where domain-general predictors had paths to all three outcomes; reading predictors had paths only to decoding and word recognition; and math predictor had a path only to math. In Model 1, learning potential factor was considered domain-general predictor and had paths to all three outcomes. It was termed as the domain-general model because I hypothesized that learning potential in one domain is generalizable to other domains. That is, the underlying mechanism of learning to decode can account not only for the concurrent decoding and word recognition skills but also for arithmetic performance. Model 1 served as the base model for determining the utility of DA in explaining academic outcomes. In Model 2, learning potential factor had paths to reading outcomes, decoding and word recognition, given the hypothesis that learning potential measured by DA is domain-specific and that learning potential for decoding is unnecessary for explaining individual differences in arithmetic. Thus, Model 2 was termed as the domain-specific model. In Model 3, learning potential factor had a path only to decoding based on the possibility that DA of decoding may measure learning potential that is specific to decoding. In this model, I assumed that learning potential for decoding is unnecessary for explaining individual differences in word recognition and arithmetic. Thus, Model 3 was termed as the task-specific model. In Model 4, learning potential did not have paths to any of the outcomes, and I assumed that DA does not have incremental validity. This model posited that learning potential for decoding does not give additional information for explaining individual differences associated with basic academic outcomes beyond what can be gathered from traditional assessments. Model 4 was termed as the null model. Summaries of the model fit results are presented in Table 6.

Research question 2 and 3 were answered by sequentially comparing three structural models (Table 4). First, Model 1 was compared with Model 4 to address whether DA is

necessary for explaining academic outcomes in general, in the presence of all other academic predictors. Results indicated that Model 1 represents the data structure better than Model 4, $\Delta \chi^2$ (3) = 12.14, $p < .05$, suggesting DA has incremental validity in explaining academic outcomes. Because Model 1 fit statistics provide evidence for the incremental validity of DA as a predictor of academic skill, alternative models were run to further examine the specific relationship of DA with each outcome. This allowed me to pinpoint the domain in which DA has predictive value. I fixed the path coefficient of the learning potential factor to arithmetic (Model 2), word recognition (Model 3), cumulatively and sequentially. To examine whether DA has incremental validity within the domain of reading but not in arithmetic, Model 1 was compared with Model 2. Model 2 did not result in poorer fit to the data structure than Model 1, $\Delta \chi^2$ (1) = .80, $p > .05$. Thus, Model 2 was selected given its parsimony. Model 2 was then compared with Model 3 to investigate whether DA was only necessary for explaining decoding. Model 3 did not fit to the data significantly worse than Model 1, χ^2 (1) = .91, $p > .05$, indicating Model 3 would be the model that best represents the data in the most parsimonious way.

Table 4.
Fit Indices for the Competing Models for Research Question 2 and 3.

Model	*df*	χ^2	SCF	*p*	CFI	TLI	SRMR
Measurement Models							
Predictor Model	60	129.726	1.014	.000	.909	.861	.063
Outcome Model	11	55.697	.989	.000	.951	.919	.037
Structural Models							
Model 1: Domain-general model	171	330.642	.990	.000	.906	.876	.062
Model 2: Domain-specific model	172	331.244	.991	.000	.906	.877	.063
Model 3: Task-specific model	173	332.153	.991	.000	.906	.878	.062
Model 4: Null model	174	343.146	.991	.000	.900	.871	.064

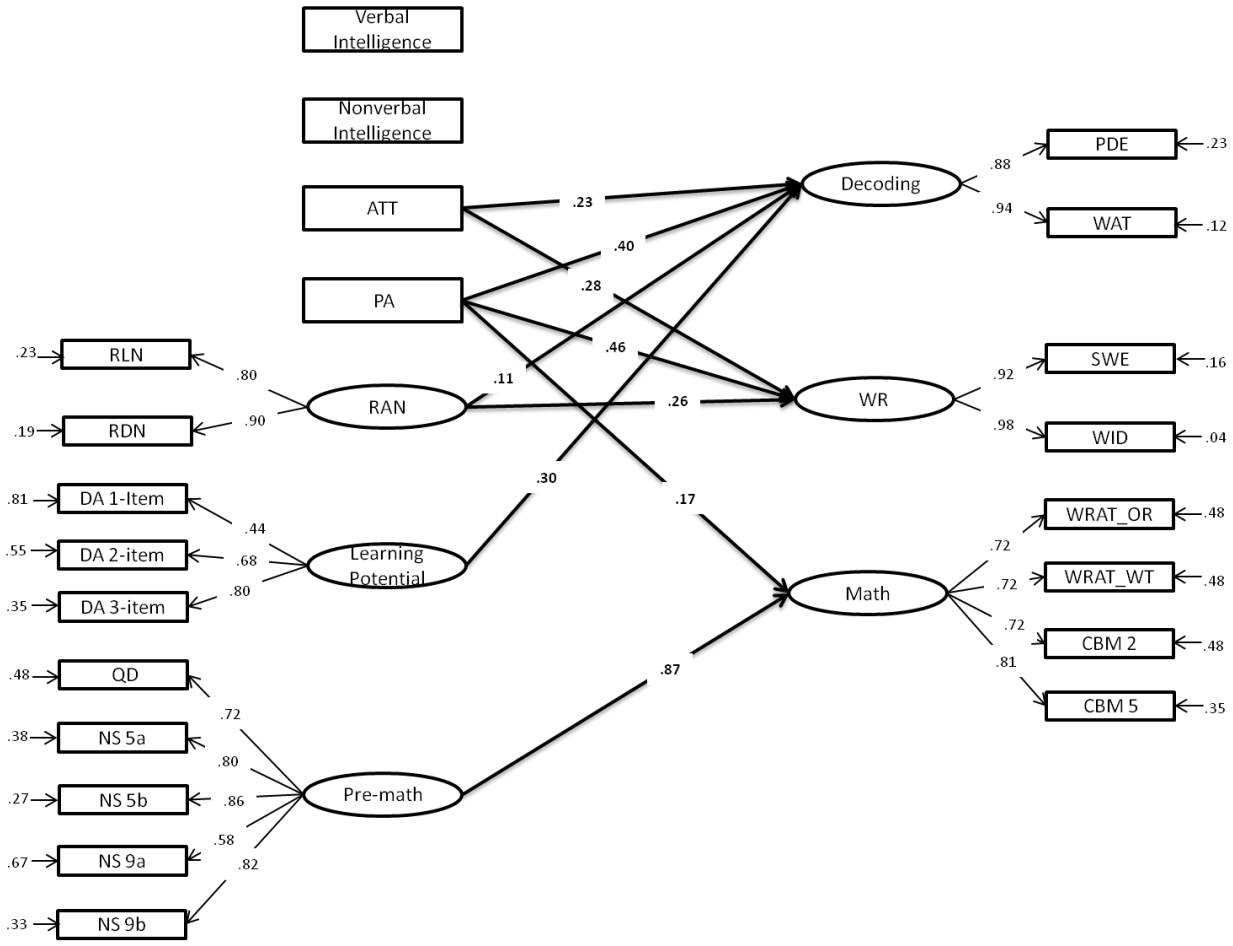


Figure 5. Final model for research questions 2 and 3. Only the significant coefficients are depicted in the figure. Coefficients were all significant at .05 alpha level. WR = Word Recognition; Math = arithmetic

CHAPTER V

DISCUSSION

General Discussion

The primary purposes of the present study were (a) to examine whether students' learning potential for decoding, measured with DA, is a separable dimension from general learning aptitude and static decoding level (construct validity), (b) to investigate whether DA has additive value in explaining students' basic academic skills (incremental validity within domain), and (c) to examine the possibility of using DA to predict math performance as well (incremental validity across domain). Despite growing interest in DA for its utility as an early screening tool for academic difficulties, there have been relatively few attempts to empirically investigate the nature of the construct that DA measures. There are multiple possibilities for what DA could tell us about individual differences in learning in comparison to other traditional static measures. First, it is possible that DA may not offer anything new. Second, DA could measure students' learning potential that represents a unique source of individual differences in learning rather than already learned knowledge and acquired skills. If this is the case, important theoretical and practical questions arise about the utility of the DA measure. For instance, does DA explain students' academic skill beyond which that can be explained by other predictors of the academic skill? Can learning potential measured with DA be generalized across domains and have predictive validity across domains? In addressing these questions, the present study extended prior reading DA literature in three ways: (a) by taking a confirmatory approach to test the

construct validity of DA; (b) by including a broad range of competing predictors to test the incremental validity of DA; and (c) by incorporating cross-domain academic criterion (arithmetic) to test the generalizability of the learning potential factor.

Construct validity of DA

First, given the different theoretical orientations between DA and static assessments, confirmatory factor analyses were conducted to empirically test the distinctiveness of the construct measured via DA. Two other important factors considered in the nomothetic span of learning potential for decoding were static decoding skills and general learning aptitude (i.e., general intelligence). Prior DA studies in reading were limited in testing the theory because of their exploratory approach (Fuchs et al., 2011) or of their lack of appropriate model comparisons (Swanson & Howard, 2005). Comparing competing theoretical models with confirmatory factor analyses allowed me to picture relationships among the measures of general cognitive abilities, decoding, and DA. Results of the present study indicate that DA is a measure of a construct distinct from general cognitive ability and decoding. This finding coincides with others' results who found models that have DA as a separate construct from the pretest performance on DA (Day et al., 1997) or from other static assessments (Fuchs et al., 2008) provided better fit to the data than the models that collapsed DA with static measures.

The distinction between static decoding measures and DA may have risen from the particular focus of each assessment. Static measures focus on assessing past learning whereas DA focuses on measuring future learning. Beginning readers' phonological processing skills (including decoding) develop rapidly with reading instruction (Wagner et al., 1993). Thus, a statically assessed decoding skill is heavily influenced by environmental factors such as educational opportunities in school or at home. On the contrary, the DA of decoding attempted to

eliminate the possibility of such influences by teaching students to decode using new orthographic system as part of the assessment. This allowed me to level the playing field and assess students' learning potential for decoding that is relatively independent from prior learning experiences. This claim can be, in part, supported by the fact that I found significant differences in schools for all of the indicators of decoding (F (5, 106) = 3.65, $p < .05$ for untimed decoding, F (5, 106) = 5.18, $p < .05$ for timed decoding, and F (5, 106) = 5.04, $p < .05$ for DA word list). However, no significant school differences were found in DA. Also supporting the claim that the DA may be relatively independent of prior reading skills comes from correlation results. Correlations between the individual words in DA word list and the corresponding items in DA mastery test ranged between .18 - .56. This suggests that the ability to read a certain English word had small to moderate relationships with students' learning potential to read that word using novel symbols, which had the same sound of that word.

The distinction between traditional intelligence assessments and DA may have been caused by two possible explanations. First, traditional assessments may measure cognitive abilities independently performed whereas DA may measure response to instructional prompts that are qualitatively different from each other. In fact, prior research provides evidence that DA of decoding is predictive of students' growth (slope) in response to small group reading instruction when controlling for verbal intelligence (Cho et al., in press). Another reason why traditional intelligence assessments and DA were separated may be attributed to domain specificity of DA. Traditional intelligence assessments tap domain general learning aptitude, whereas DA of decoding taps learning potential specific to its task domain. The plausibility of this explanation is addressed in the third research question, which will be discussed later.

Incremental validity of DA

Because DA used in this study provided *unique* information of individual differences in learning to read, the next step was to ask whether this information gathered from DA, learning potential, is *important* in explaining students' academic performance even when all other predictors are controlled. This relates to the second contribution of this study to the extant DA literature. In the present study, incremental validity of DA was examined using a more stringent criterion than previous studies by including a wide range of potentially important predictors of word reading. I included domain general cognitive abilities, both verbal and nonverbal intelligence, and the teacher's rating of attentive behavior in addition to phonological awareness and rapid automatized naming for explaining decoding and word recognition.

Testing the incremental validity of DA as a measure is a practically important issue. Phonological awareness and rapid automatized naming of letters/digits are considered two important factors of dyslexia (e.g., Wolf & Bowers, 1999), and they are the most common and strongest measures that were used to predict beginning reading development (e.g., Compton, 2000). If DA provides additional information to what commonly used predictors of reading can offer, DA may be used as a helpful tool for practitioners to correctly identify students with reading difficulties.

Results suggest that DA indeed provides important information of individual differences in decoding. Students' potential for learning to decode provided more comprehensive picture of students' decoding skills in addition to students' phonological skills and attentive behaviors. Results suggest students' cognitive ability is best understood as a process. Instead of focusing on the end product of students' development, students' cognitive abilities can be fully understood by considering his/her continually developing abilities mediated by interaction with others.

Therefore, how well students learned to decode in response to instruction provided a more comprehensive picture of their concurrent decoding skills.

Although DA showed significant positive relationship with word recognition, DA was not necessary for explaining word recognition in the presence of competing predictors. This may be due to the fact that the development of word recognition skills involves more than learning grapheme-phoneme relations and how to use these relations to decode words. In fact, English includes numerous high-frequency words that are not decodable and can only be read from memory. Another reason why DA failed to have incremental validity for word recognition is due to the limited number of vowel sounds used in DA. Decoding tasks used in DA included only one vowel sound, /a/, when the understanding of the variants of vowel sounds is critical for developing more complete representations of sight words in memory (Ehri, 1999).

This study also contributed to the DA literature in terms of the breadth of the outcomes. Prior studies that had both reading and math outcomes used DA tasks that are more domain-general in nature measuring basic intelligence or working memory (Swanson & Howard, 2005). Studies that had DA tasks in a specific academic domain did not include outcomes outside of that domain (Bridges & Catts, 2011; Cho et al., in press; Fuchs et al., 2008; Fuchs et al., 2011; Seethaler et al., 2011; Spector, 1992). I used domain-specific DA tasks (decoding) while including arithmetic outcome to examine whether learning potential is generalizable across domains. Including arithmetic outcomes served two purposes in this study. First, it allowed me to test whether DA measures domain-general or domain-specific learning potential. In Grigorenko's (2009) theoretical paper comparing DA and response to intervention (RTI), she pointed out that domain generality of learning potential is an unsettled issue whereas RTI measures students' responsiveness that is domain-specific. If DA and RTI are truly parallel

concepts, DA may also measure response to instructional prompts specific to the domain of the task. Results from this study suggest that DA measures learning potential that is not generalizable across domains. Second, given the fact that DA is highlighted as a useful screening tool that can supplement existing static measures, examining incremental validity of DA across domains could guide practitioners whether different DA should be used in screening for different outcomes. The present results suggest not using a single DA for documenting students' responsiveness across different domains.

Limitation and Future Direction

Although the present study extended existing DA research in important ways, the results must be interpreted in light of the three study limitations. Two of the limitations deal with statistical concerns and the other is related to theoretical and practical issues. Each limitation suggests areas for future research.

The first statistical concern is small sample size. Results should be interpreted with caution because of the small sample size. It may be noted, however, that even with a relatively small sample size and with limited power to detect differences between models, I did detect significant differences using the chi-square difference tests across competing models. The other statistical concern is the use of manifest variables. Because the initially proposed measurement model did not represent the data structure well enough, the use of manifest variables in the models was unavoidable. Although the measures entered as manifest variables had high reliabilities, path coefficients can be over- or under- estimated when fallible measures are used in the analyses. Thus, future research on DA using structural equation modeling may benefit from

incorporating larger samples and obtaining multiple measures for use in modeling to get more stable estimates.

Results are limited to DA's concurrent prediction of academic skills. I did not examine the DA's predictive validity in forecasting future performances. One of the advantages of DA over traditional assessments is that DA focuses on measuring what students can do in the future instead of what students can do at the time of testing. Thus, in theory, DA should have additive value in predicting students' future performance. Moreover, its predictive value may be greater when explaining future performances than concurrent ones because DA is a more sensitive measure for growth than for the present level of performance.

Depending on when the outcomes are measured, different results could occur. Interestingly, Seethaler et al. (2011) found that DA of math using balancing equation tasks had stronger incremental validity for predicting future outcomes that were more distal to the DA tasks (word problem) than for more proximal task (computation). This is somewhat contrary to the present results, which found incremental validity of DA for outcomes that are proximal to DA task (decoding) but not for distal to DA tasks (word recognition and arithmetic). In the present study, DA had incremental validity only when the outcome was strictly aligned to the task demand of DA.

There are two important differences between Seethaler et al. (2011) and the present study that could account for the different patterns of results. First, the domain is different. Seethaler and her colleagues used math DA, whereas the present study used reading DA. Second, Seethaler et al. used DA to predict future performance, and the present study focused on concurrent prediction. Thus, DA may work better when the outcome closely aligns with the DA tasks than distal outcome when predicting concurrent outcomes. On the other hand, DA may

work better when the outcome is distal to the DA tasks when predicting future outcomes. It is noteworthy that learning and transfer is the key concept of graduated prompts approach (Campione & Brown, 1990). Students who can master the decoding skills only with implicit instruction prompts have better transfer ability than those who need more explicit instructional prompts. Given this premise, I would expect to find such patterns, were I to examine predictive validity. For example, DA of decoding may have incremental validity for predicting future word recognition for the following reason. According to the self-teaching hypothesis (Share, 1995), decoding skills transfers to word recognition skills. In other words, repeated practice in decoding is critical in building accurate representation of word's orthographic unit and in strengthening of phonological information associated to the letters. Substantial practice time is also required for decoding skills to transfer to word recognition skills. Thus, if the present DA of decoding measures students' learning potential for decoding and quantifies ability to transfer, it might predict how well students can transfer their decoding skills to word recognition skills. If I had allowed time for such transfer to occur and predicted future word recognition skills, as in prior studies (Bridges & Catts, 2011; Cho et al., in press; Fuchs et al., 2011; Spector, 1992), I could find incremental validity of DA for predicting word recognition development. Thus, the next step in this line of research is to examine the predictive validity of DA.

Conclusion

Overall findings of this study provide insight into cognitive processes involved in first graders' reading and math performance. Particularly, this research supports the existence of Vygotsky's notion of ZPD by providing empirical evidence that there are differences between what students can do independently and what students can accomplish given assistance (learning

potential). In addition, DA of decoding provided a more complete picture of students' decoding skills beyond that which conventional assessment can provide.

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: Level 1-5 (PAL)**

Now 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? 山

This says /p/. What sound? 人

This says /f/. What sound? 魚

This says /s/. What sound? 日

This says /t/. What sound? 占

This says /a/. What sound? 石

Mastery Test DA1: 1-5

Tell me the sounds of these letters. Starting from top to bottom, 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 key words for each funny letter.

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

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

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

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

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

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

Mastery Test DA1: 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 DA1:7

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 /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 /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 /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 /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 /t/ sound as in top? Because it came from the shape of a top, this says /t/ as in top. What sound? What word?

This is a apple. And this is an easy way of drawing a apple. Now, do you see why this makes /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 DA1:8

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 DA1:9

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

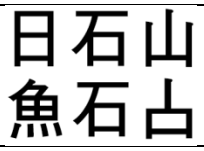
DA 2:Blending Sounds (CVC)

- **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 2-1
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.

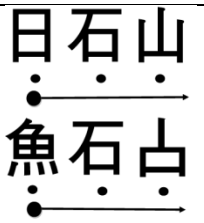
Mastery Test DA 2-2
Read these words to me. (No feedback is provided)

- **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.



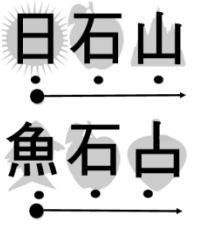
Mastery Test DA 2-3
Read these words to me. (No feedback is provided)

- **DA2: 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.

/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.



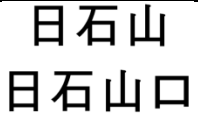
Mastery Test DA 2-4
Read these words to me. (No feedback is provided)

DA 3: Rule-based learning (CVCe)

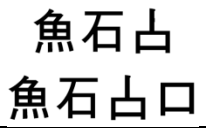
- **DA 3: Level 1**

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)

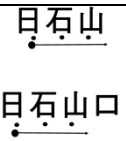


Mastery Test DA 3-1
Read these words to me. (No feedback is provided)

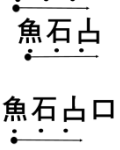
- **DA 3: Level 2**

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)

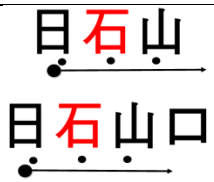


Mastery Test DA 3-2
Read these words to me. (No feedback is provided)

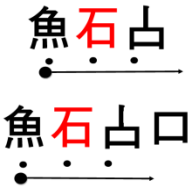
- **DA 3: Level 3**

I am going to 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 /a/ sound as in apple. But this letter in same, says /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 /a/ sound as in apple. But this letter in fate, says /ā/ sound, as in apricot.

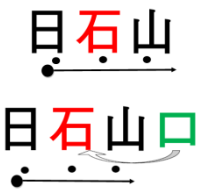


Mastery Test DA 3-3
Read these words to me. (No feedback is provided)

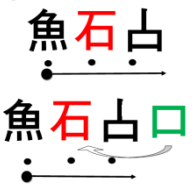
- **DA 3: Level 4**

Now, I will tell you why this letter makes different sounds.

This last funny letter in same does not have a sound. But, 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).



This last funny letter in same does not have a sound. But, 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).

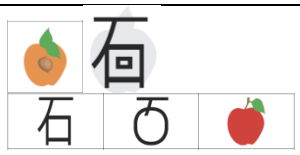


Mastery Test: DA 3-4
Read these words to me. (No feedback is provided)

DA 3: Level 5

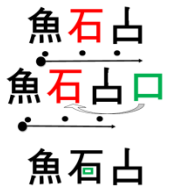
Now, I will give you the key word and picture to help you remember.

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 (as in apricot). 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 s. ā (as in apricot). m. same.
Let's try.
s.a.m. sam. Your turn.
s. ā.m. same. Your turn.



Mastery Test DA 3-5

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

REFERENCES

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin.. *Psychological bulletin*, *112*(3), 400.

Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology, 76,* 610-622.

Bransford, J. C., Delclos, J. R., Vye, N. J., Burns, M., & Hasselbring, T. S. (1987). State of the art and future directions. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 479–496). New York: Guil-ford Press.

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.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions, 154,* 136-136.

Budoff, M. (1967). Learning potential among institutionalized young adult retardates. *American Journal of Mental Deficiency, 72,* 404–411.

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*(3), 403.

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. (1987). Linking dynamic assessment with school achievement. In C. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82–115). New York: Guilford Press.

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.

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.

Carroll., J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

Chard, D. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention, 30* (2), 3–14.

Cho, E., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (in press). Examining the Predictive Validity of a Dynamic Assessment of Decoding to Forecast Response to Tier 2 Intervention. *Journal of Learning Disabilities.*

Clarke, B., Baker, S., Smolkowsi, K., & Chard, D. J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. Remedial and Special Education, 29, 46–57.

Compton, D. L. (2000). Modeling the growth of decoding skills in first-grade children. *Scientific Studies of Reading, 4,* 219-258.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological methods,* 1, 16-27.

Daniel, M. H. (1997). Intelligence testing: Status and trends. *American Psychologist, 52*, 1038-1045.

Day, J. D., Engelhardt, J. L., Maxwell, S. E., & Bolig, E. E. (1997).Comparison of static and dynamic assessment procedures and their relation to independent performance. *Journal of Educational Psychology, 89(?),* 358–368.

De Smedt, B., Taylor, J., Archibald, L., & Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills*? Developmental science, 13*, 508-520.

Ehri, L.C. (1999). Word reading by sight and by analogy in beginning readers. In C. Hulme & R.M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 87-111). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179–197.

Feuerstein, R. (1979). *The dynamic assessment of retarded performers. The Learning Potential Assessment Device, theory, instruments, and techniques.* Baltimore: University Park Press.

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, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*, *97*, 493–513.

Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Craddock, C. F., & Hamlett, C. L. (2008). Dynamic assessment of algebraic learning in predicting third graders' development of mathematical problem solving. *Journal of Educational Psychology; Journal of Educational Psychology, 100*, 829-850.

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1990). *Monitoring basic skills progress.* Austin, TX: Pro-Ed.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool the number sets test. *Journal of Psychoeducational Assessment, 27*(3), 265-279.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*, 79-132.

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.

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15*(4), 456.

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical

computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology*, *79*, 192-227.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, *3*, 424.

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.

Kelley, E. L. (1927).*Interpretation of educational measurements*.Yonkers, NY: World

Kline, R.B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). NY: Guilford Press.

Lidz, C. S. (1987). *Dynamic assessment: An interactional approach to evaluating learning potential.* NY: Guilford Press.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904)'"General intelligence,'objectively determined and measured". *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology*, *86*, 96.

Minick, N. (1987). Implications of Vygotsky's theory for dynamic assessment. In C.S. Lidz (Ed.), *Dynamic Assessment: An interactional approach to evaluating learning potential* (pp.116-140). New York: Guilford.

Muthén, L. K., & Muthén, B. O. (1998-2010). Mplus user's guide: The comprehensive modeling program for applied researchers. Los Angeles: Muthén & Muthén.

Research Institute on Progress Monitoring. (2009). Early numeracy indicators (*Number Identification, Quantity Discrimination, Missing Number, Mixed Numeracy*). Retrieved July 20, 2012, from http://www.progressmonitoring.org/pdf/TREarlymath6.pdf

Ryan, J. J., Carruthers, C. A., Miller, L. J., Souheaver, G. T., Gontkovsky, S. T., & Zehr, M. D.(2003). Exploratory factor analysis of the Wechsler Abbreviated Scale of Intelligence (WASI) in adult standardization and clinical samples. *Applied Neuropsychology, 10,* 252–256

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *1988 American Statistical Association Proceedings of the Business and Economics Section* (pp. 308-313). Alexandrai VA: American Statistical Association.

Sanders, C. E., Lubinski, D., & Benbow, C. P. (1995). Does the Defining Issues Test measure psychological phenomena distinct from verbal ability?: An examination of Lykken's query. *Journal of Personality and Social Psychology, 69,* 498–504.

Seethaler, P. M., Fuchs, L. S., Fuchs, D., & Compton, D. L. (2011). Predicting first graders' development of calculation versus word-problem performance: The role of dynamic assessment. *Journal of Educational Psychology, 104*(1), 224.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

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

StataCorp. (2009). *Stata Release 11. Statistical Software.* College Station (TX): StataCorp LP

Sternberg, R. J. (1996). *Successful intelligence*. New York: Simon & Schuster.

Sternberg, R. J., & Detterman, D. K. (1986). *What is intelligence?: Contemporary viewpoints on its nature and definition*. Ablex Publishing Corporation.

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., & Howard, C. B. (2005). Children with reading disabilities: Does dynamic assessment help in the classification? *Learning Disability Quarterly, 28,* 17-34.

Swanson, J., Schuck, S., Mann, M., Carlson, C., Hartman, K., Sergeant, J., et al. (2004). Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and the SWAN rating scales. Retrieved July 20, 2012, from http://www.adhd.net/SNAP_SWAN.pdf

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

Tzuriel, D., & Haywood, H. C. (1992). The development of interactive dynamic approaches for assessment of learning potential. In H.C. Haywood & D. Tzuriel (Eds.), *Interactive assessment* (pp. 3-37). New York: Springer-Verlag.

Vygotsky, L. S. (1962). *Thought and language.* Cambridge, MA: MIT Press (Original work published 1934).

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

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85,* 83.

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

Wilkinson, G. S. (1993). *Wide Range Achievement Test 3.* Wilmington, DE: Wide Range.

Wolf, M. (1986). Rapid alternating stimulus naming in the developmental dyslexias. *Brain and Language, 27,* 360-379.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of educational psychology*, *91*, 415.

Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests—Revised.* Circle Pines, MN:American Guidance Service.