

THE USE OF A DYNAMIC SCREENING OF PHONOLOGICAL AWARENESS
TO PREDICT READING RISK FOR KINDERGARTEN STUDENTS

C2009
by Melinda Sittner Bridges

Submitted to the Department of Speech-Language-Hearing: Sciences and Disorders
and the Faculty of the Graduate School of the University of Kansas
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

B.S., University of Iowa, 1994

M.S., University of Nebraska-Lincoln, 1996

Hugh W. Catts, Ph.D., Chairperson*

Diane Frome Loeb, Ph.D.

Committee Members

Diane C. Nielsen, Ph.D.

Mabel L. Rice, Ph.D.

Holly L. Storkel, Ph.D.

Date defended: _____

The Dissertation Committee for Melinda Sittner Bridges certifies

that this is the approved version of the following dissertation:

THE USE OF A DYNAMIC SCREENING OF PHONOLOGICAL AWARENESS
TO PREDICT READING RISK FOR KINDERGARTEN STUDENTS

Committee:

Hugh W. Catts, Ph.D., Chairperson

Diane Frome Loeb, Ph.D.

Diane C. Nielsen, Ph.D.

Mabel L. Rice, Ph.D.

Holly L. Storkel, Ph.D.

Date approved: _____

ABSTRACT

Response to intervention has been proposed as a framework for early identification and intervention. In such a framework, all students receive periodic screening (i.e., universal screening) for risk for reading disabilities; those identified as at risk are provided with supplemental intervention. In order for such an approach to be successful, universal screening measures must correctly identify students truly at risk. The purpose of this study was to investigate the usefulness and predictive validity of a dynamic screening of phonological awareness. In Study I, the dynamic screening measure was administered to students at the beginning of kindergarten. The results demonstrated that the dynamic screening measure can be administered by school personnel with high reliability. Additionally, the distributional characteristics of the dynamic screening of phonological awareness were compared to those of other phonological awareness measures. Although the dynamic screening measure had a low skewness statistic, many students scored a zero on this measure. However, the other phonological awareness measures showed more of a floor effect.

In Study II, a revised dynamic screening measure was administered to two samples of kindergarten students. Results showed the distribution of the dynamic screening measure did not have a floor effect. The predictive validity of the dynamic screening measure was compared to a static measure containing the same test items. The results indicated that the dynamic screening measure significantly improved the prediction of reading outcomes over and above the static measure, suggesting that the dynamic nature of the former contributed to the prediction accuracy. The predictive

validity of the dynamic screening measure was also compared to a commonly used phonological awareness screening measure. Results showed that the dynamic screening measure added significantly to the prediction of reading outcomes. Additional analyses examined the use of the dynamic screening measure as a supplemental measure. The findings demonstrated that the dynamic measure reduced the number of false positives, and in some cases, predicted reading outcomes as well as a combination of the two measures. The results of this study provide preliminary support for the usefulness of a dynamic screening of phonological awareness within an RTI framework for kindergarten students.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Hugh Catts, for his guidance throughout my graduate school career. I truly appreciate his patience, wisdom, and sense of humor. He has been a true mentor and has always encouraged my need to link research to practice. I would also like to thank the members of my comprehensive exams and/or dissertation committees: Drs. Diane Frome Loeb, Diane Nielsen, Mabel Rice, Holly Storkel, and Marc Fey. Each individual had a positive influence on my graduate studies that led to this dissertation, and for that I am grateful.

I would also like to thank the school districts who graciously participated in this research: Lawrence public schools, Van Meter Elementary, Pembroke Elementary, Menlo Elementary, and I-35 Elementary. This work would not have been possible without their involvement.

Thanks to Dr. Kandace Fleming and Angela Miller for help with statistical analyses. I would also like to thank Chris Lorenzen for helping me draft an initial version of the forms and artwork used in the pilot work for this dissertation.

This work was financially supported in part by a National Institutes of Health Institutional Traineeship (T32 DC000052).

My fellow students/friends in the Language and Reading Disorders Laboratory have been a constant source of knowledge, support, and encouragement. I look forward to continued professional and personal relationships with each of them.

I have so many friends who have supported me during this process. My graduate school friends have critiqued papers, listened to presentations, celebrated

with me during the good times, and commiserated with me when times were tough. The transition back to graduate school would have been incredibly difficult without their support. My friends on the “outside” reminded me that I should still participate in “real life.” Although I will never be able to repay them, I appreciate their patience, enthusiasm, and continued interest in my research project.

My amazing family is responsible for the initiation of my graduate school career. Our family outings to the library instilled in me a love of books and learning that led me to this point. My parents believed that I could do or be whatever I wanted, and they were called upon to remind me of that more times than I can count.

Finally, my fabulous husband is responsible for the completion of my graduate school career. I thank him for his patience over the years as well as for his interest in my dissertation research. The look of pride on his face when he talks about “my wife in graduate school” has carried me through many long nights. He has been incredibly unselfish while I have pursued this dream. For always and forever, I will be grateful to him.

TABLE OF CONTENTS

Acceptance Page.....	2
Abstract.....	3
Acknowledgements.....	5
Table of Contents.....	7
List of Tables.....	9
Chapter 1: Introduction.....	11
Response to intervention to identify reading disabilities.....	12
Prediction of reading disabilities.....	14
Evaluation of screening measures.....	17
Multivariate screening.....	19
Investigating more efficient screenings.....	21
Dynamic assessment.....	23
Curriculum-based dynamic assessment.....	28
Dynamic assessment as a screening for reading disabilities.....	29
Limitations of past research on dynamic assessment.....	33
Overview of the present study.....	35
Chapter 2: Study I.....	37
Method.....	37
Results and discussion.....	42
Chapter 3: Study II.....	47
Methods.....	48

Results.....	52
Chapter 4: General Discussion.....	79
A dynamic approach to universal screening.....	84
Limitations of the current research study.....	88
Considerations for future research.....	89
References.....	91
Figures.....	107
Appendix.....	113

LIST OF TABLES

- TABLE 2.1 Descriptive statistics for predictor variables
- TABLE 3.1 Administration schedule for Sample 1 and Sample 2
- TABLE 3.2 Descriptive statistics for Sample 1 (N=90)
- TABLE 3.3 Descriptive statistics for the full Sample 2 (N=161)
- TABLE 3.4 Descriptive statistics for the reduced Sample 2 (N=96)
- TABLE 3.5 Correlations between predictors and outcome measures in Sample 1 (N=90)
- TABLE 3.6 Hierarchical linear regression analyses predicting Word Identification and Word Attack in Sample 1 (N=90)
- TABLE 3.7 Logistic regression analyses for predicting Word Identification and Word Attack performance in Sample 1 (N=90)
- TABLE 3.8 Correlations between predictors and outcome measures
- TABLE 3.9 Hierarchical regression analyses for predicting Word Identification and Word Attack in Sample 1 (N=90)
- TABLE 3.10 Hierarchical regression analyses for predicting Word Identification and Word Attack in the full Sample 2 (N=161)
- TABLE 3.11 Hierarchical regression analyses for predicting Word Identification and Word Attack in the reduced Sample 2 (N=96)
- TABLE 3.12 Classification indices for the logistic regression analyses for Sample 1 (N=90)
- TABLE 3.13 Classification indices for the logistic regression analyses for the full

Sample 2 (N=161)

- TABLE 3.14 Classification indices for the logistic regression analyses for the reduced Sample 2 (N=96)
- TABLE 3.15 Classification indices across models utilizing the ISF alone and the ISF with the DSPA-R as a supplemental screening measure in Sample 1 (N=90)
- TABLE 3.16 Classification indices across models utilizing the ISF alone and the ISF with the DSPA-R as a supplemental screening measure in the full Sample 2 (N=161)
- TABLE 3.17 Classification indices across models utilizing the ISF alone and the ISF with the DSPA-R as a supplemental screening measure in the reduced Sample 2 (N=96)
- TABLE 3.18 Classification indices in two-step process maintaining sensitivity above .90 and reducing false positives in Sample 1 (N=90)
- TABLE 3.19 Classification indices in two-step process maintaining sensitivity above .90 and reducing false positives in the full Sample 2 (N=161)
- TABLE 3.20 Classification indices in two-step process maintaining sensitivity above .90 and reducing false positives in the reduced Sample 2 (N=96)
- TABLE 3.21 Specificity for the supplemental model, the combined model, the DSPA-R alone, and the ISF alone when sensitivity is held constant to that found with the supplemental model

CHAPTER I: INTRODUCTION

Poor reading achievement has been a longstanding concern in the United States. Currently, a substantial number of students across the country do not read well enough to perform successfully in school (National Reading Panel, 2000). In 2007, 33% of fourth graders read below the “basic” level on the National Assessment of Educational Progress (NAEP) reading test (National Center for Education Statistics, 2007), and this low number has remained relatively constant over the last 25 years. This percentage is significant given research indicating that a poor reader in elementary school is likely to be a poor reader in adolescence and adulthood (for a review, see Scarborough, 1998; Juel, 1988).

One significant factor associated with improved outcomes of children at risk for reading disabilities (RD) is early identification. When young children at risk for RD are identified and provided with explicit and systematic reading instruction based on concepts such as phonological awareness, fluency, and oral language, the incidence of later reading disabilities is greatly reduced (Denton & Mathes, 2003; Scanlon & Vellutino, 1996; Vellutino, Scanlon, Small, & Fanuele, 2006). In fact, converging research indicates that when intervention begins in kindergarten, it can result in a substantial portion of students at risk for RD achieving normal reading proficiency in first grade and beyond (Cavanaugh, Kim, Wanzek, & Vaughn, 2004; O’Conner, Fulmer, Harty, & Bell, 2005; Scanlon, Vellutino, Small, Fanuele, & Sweeny, 2005; Simmons et al., 2008; Vellutino et al., 2006). For example, Simmons and colleagues (2008) showed that 94% of students deemed at risk attained normal

reading proficiency in third grade if they had received intervention in kindergarten. However, the success of early prevention and intervention procedures hinges on the ability to identify students who are at risk prior to the onset of RD. As Torgesen (1998, p. 34) stated, “in order to efficiently remediate, we must identify the right children at the right time.”

Response to Intervention to Identify Reading Disabilities

It has been suggested that the assessment of students’ response to high-quality instruction is a viable method for identifying RD (Fletcher, Coulter, Reschly, & Vaughn, 2004; Fuchs & Fuchs, 2005; Haager et al., 2007; Vaughn & Fuchs, 2003). Recently, such an approach has been added as an identification option in the reauthorization of the Individuals with Disabilities Education Improvement Act (IDEA, 2004). According to this option, students may be identified as having RD if their response to scientifically-based instruction, including targeted intervention, is substantially below their peers. This identification and prevention approach has been formally called *Response to Intervention* (RTI; Denton & Mathes, 2003; Fuchs & Fuchs, 1997; Gresham, 2002; Haager et al., 2007; Vellutino, Scanlon, & Zhang, 2008). Typically, RTI consists of three tiers of sequentially ordered instruction/intervention. In Tier 1, all students receive high-quality literacy instruction within a general education setting. Periodic universal screening is used to measure response to this classroom instruction. Those students identified as at risk based on universal screenings then participate in more intensive, small-group intervention (Tier 2). Students who fail to respond to this additional intervention are considered to be

truly at risk for RD; at this time, students are often tested more extensively and provided with even more individualized intervention (Tier 3).

Despite the promise of RTI to improve the identification of RD, relatively few studies have specifically examined the effectiveness of this approach for *early* identification (Compton et al., 2006; VanDerHeyden, Witt, & Gilbertson, 2007; Vellutino et al., 1996; Vellutino et al., 2008). Compton et al. (2006) showed that a combination of brief literacy assessments and progress monitoring in first grade was an accurate predictor of RD in second grade. Further results also suggested that differential response to short-term intervention during first grade may significantly add to the prediction of reading problems (Fuchs, Compton, Fuchs, Bryant, & Davis, 2008). In another longitudinal study, Vellutino and colleagues (Vellutino et al., 2008) provided evidence that an RTI identification approach is a viable method of identifying risk for RD in kindergarten. A universal screening was administered to all students at the beginning of kindergarten. Half of those students received small group intervention during their kindergarten year, and the other half received whatever services were offered in their school (i.e., “business as usual”). A series of logistic regression analyses were carried out to assess the accuracy with which measures of response to kindergarten intervention would predict end of first grade reading. Results indicated that measures of growth in early literacy skills, used as indices of response to kindergarten intervention, were excellent predictors of end of first grade word reading over and above more traditional measures, such as letter identification and phonological awareness. These results provide evidence that an RTI approach to RD

classification can be an effective method of identifying students who at risk for RD and therefore require more individualized special education. Furthermore, it seems an appropriate identification process to be used with students entering kindergarten, which would allow for supplemental intervention to be provided in a timely fashion.

Prediction of Reading Disabilities

Converging evidence indicates that RTI is a promising method of early identification, but its success depends in part on the ability of an early screening measure to identify those students truly are at risk for RD. One of the difficulties associated with early identification is that the earlier a screening measure is administered to students, the less “valid and potent” a predictor it becomes (Gersten & Dimino, 2006); this is most likely because early screening measures can’t directly assess word reading. However, given the probability that, without intervention, a poor reader will remain poor across time, screening measures are most optimally administered to students in the early stages of formal education. Choosing an appropriate screening measure for a skill that is not yet present (i.e., reading) is difficult but critical in an early identification framework. Thus, a sensitive screening measure must target those pre-reading skills that, if impaired, would most likely signal a later reading disability.

A longstanding line of research has attempted to identify which pre-reading skills best predict reading outcomes (e.g., Badian, 1994; Catts et al., 2001; Compton et al., 2006; Elbro, Borstrom, & Petersen, 1998; Scarborough, 1998). Although many early literacy measures have been used to identify risk for RD, three of the most

widely investigated predictors are letter name/sound knowledge, rapid automatic naming (RAN), and phonological awareness.

A large body of evidence has shown assessments of letter/sound knowledge to be predictive of reading achievement (e.g., Catts et al., 2001; Elbro et al., 1998; Pennington & Lefly, 2001; Scanlon & Vellutino, 1996; Scarborough, 1998; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Whitehurst & Lonigan, 1998). Children who demonstrate problems in learning letter names/sounds often develop RD. One potential problem with using a screening measure of letter knowledge with students at the beginning of kindergarten is the potential for a floor effect, most likely due to a students' limited literacy experience. Thus, a student unexposed to explicit letter/sound instruction prior to starting kindergarten might score at very low levels on a letter naming test initially, but after a brief time in kindergarten, could easily score within normal limits.

Rapid automatic naming, which is defined as the ability to quickly name randomly repeated visual stimuli such as objects, letters, and digits, has also proven to be a good predictor of RD (e.g., Badian, 1993; Bowers, 1995; Felton & Brown, 1990; McBride-Chang & Manis, 1996; Schatschneider et al., 2004). Some have found the relationship between RAN and reading to vary depending on the RAN tasks used (e.g., Schatschneider et al., 2004; Savage and Fredrickson, 2005). For example, Schatschneider et al. (2004) suggested that RAN of letters was a better predictor of later reading than RAN of objects. However, RAN of letters has also been found to

over identify students at the beginning of kindergarten, particularly those that have had limited literacy exposure and/or instruction.

For the last three decades, the majority of studies investigating early identification of RD have included measures of phonological awareness (e.g., Catts et al., 2001; Mann, 1993; O'Connor & Jenkins, 1999; Schatschneider et al., 2004; Stanovich, Cunningham, & Cramer, 1984; Wagner & Torgesen, 1987). Phonological awareness refers to the ability to focus on and manipulate phonemes in spoken words (Gillon, 2004). The causal relationship between phonological awareness and later reading ability has been well established (for reviews, see Adams, 1990; Catts & Kamhi, 2005). Young children who are successful at tasks such as detecting and manipulating syllables, rimes or phonemes typically are quicker to read than those who are not; this is true even when factors such as IQ and receptive vocabulary are controlled (e.g., Bryant, MacLean, Bradley, & Crossland, 1990; Wagner & Torgesen, 1987). Children's performance on tasks of sound segmentation and sound blending in kindergarten has also been identified as a good predictor of reading ability at the end of first and second grade (e.g., Mann, 1993; Perfetti, Beck, Bell, & Hughes, 1987; Stanovich et al., 1984; Torgesen, Wagner, & Rashotte, 1994). Students with reading disabilities typically perform more poorly on tasks of phonological awareness than their peers without RD (e.g., Blachman, 1989; Catts et al., 2001; Stanovich, 1986; Vellutino & Scanlon, 1987).

Evaluation of Screening Measures

Studies examining the predictors of RD have typically used linear regression or similar statistical methods to evaluate the accuracy of prediction. However, early identification procedures require additional statistical indices to more fully explore classification accuracy. There are several indices available to evaluate a screening measure. A screening measure can be correct in two ways (Dollaghan, 2007; Jenkins, 2003). First, it can correctly identify a child as at risk (true positive). Second, it can correctly identify a child as not at risk (true negative). In the same manner, a screening measure can be incorrect in two ways. It can identify a child as at risk, when in fact the child is not (false positive). It can also identify a child as not at risk, when in fact the child is at risk (false negative).

Additionally, two other statistics can be used to gauge a screening measure's accuracy: *sensitivity* and *specificity*. Sensitivity is the percentage of poor readers correctly identified by the screening measure. Specificity is the percentage of good readers correctly identified by the predictor. As a screening measure correctly identifies a higher number of students who will have later RD, the sensitivity of the measure increases. As a screening measure correctly identifies more students who will not have RD, the specificity increases.

Many researchers have designed their screening criteria to have high sensitivity (Scanlon & Vellutino, 1996; O'Connor & Jenkins, 1999). This is important in a school setting, as the goal is to ensure that all students are receiving appropriate additional intervention when needed. However, over-identification of students as at

risk for RD means providing additional intervention to students who do not need it, causing an unnecessary strain on a school's personnel and budget. There does not seem to be a general consensus for indices of specificity and sensitivity for a screening measure. Some have suggested a level of .80 for both indices to be an acceptable level (Carran & Scott, 1992; Jansky, 1978; Kingslake, 1983). More recently, researchers have recommended the sensitivity be at least .90 to be acceptable (Jenkins, Hudson, and Johnson, 2007). Jenkins (2003) suggested that when judging the classification accuracy of a screening measure, it is best to identify a measure with adequate sensitivity and then evaluate the acceptability of the false positive rate (i.e., 1-specificity). In an RTI framework, the goal of universal screening is to have very few false negatives by using screening measures that yield true-positive rates approaching 100% (Jenkins, 2003; Jenkins & Johnson, 2008). Due to this desire to capture all students who are truly at risk, the false positive rate of early screening is often as high as 50% (Dickman, 2006).

Another way to examine the predictive ability of a screening measure is to use the ROC (receiver operating characteristic) curve. Often used in medical research when screening for a medical diagnosis, a ROC curve is a plot of true positive rate (i.e., sensitivity) versus false positive rate for each of the possible cutoff scores of the predictor. Area under the curve (AUC), an index of the area under a ROC curve, provides an overall estimate of the predictability of a measure. The AUC is an estimate of how accurately a screening measure will classify two randomly chosen

individuals, one from the poor outcome group and one from the good outcome group. Values of AUC range from .5 (i.e., chance level) to 1.0 (i.e., perfect classification).

Multivariate Screening

In an attempt to optimize the accuracy of identification, many early identification studies have employed a multivariate screening approach (e.g., Catts et al., 2001; Compton et al., 2006; Fuchs, Fuchs, & Compton, 2004; Schatschneider et al., 2004). For example, Catts et al. (2001) examined early predictors of later RD using this type of approach. A large sample of students (N=604) was given a battery of early literacy and language assessments in the spring of kindergarten, including letter naming knowledge, phonological awareness, and other oral language measures. Follow-up testing was administered in second grade, and logistic regression analyses were completed to observe the set of predictors that most uniquely predicted later reading. Results indicated that letter naming, sentence imitation, phonological awareness, and RAN, along with mother's education level, were the most significant set of predictors for reading ability in second grade. The classification accuracy rates indicated that an optimal probability level resulted in a sensitivity of 92% and a specificity of 80%. Although these numbers show good predictive validity, the entire battery was too long to be utilized as a universal screening procedure. Additionally, information related to mother's education may be difficult to obtain in a general education setting.

O'Connor and Jenkins (1999) also utilized a multivariate screening approach to early identification. Their study included three cohorts of students in kindergarten

through first grade, with screening measures administered to all students on several different occasions. The researchers reported that two static measures of phonological awareness and a static measure of RAN, administered in November of kindergarten, could differentiate students with and without RD (defined as 1.4 standard deviations below the mean score on a combination of a sight word reading subtest and word attack subtest) at the end of first grade. Results of this study revealed a high sensitivity (100%) and specificity (89%). A comparable level of accuracy was reported for a similar set of measures which were administered in October of first grade (sensitivity 100%, specificity 87%). Although this screening battery resulted in high levels of sensitivity and specificity, it took approximately half an hour to complete. The screening was also completed in November; it is possible that the classification accuracy would decrease if administered at the beginning of kindergarten.

More recently, Compton, Fuchs, Fuchs, and Bryant (2006) administered a large battery of screening tests to 206 low-functioning first graders (based on rapid naming, word identification, and teacher's judgment). Reading disability was defined on the basis of poor performance on a composite of word identification and reading comprehension tests at the end of first and second grades. Logistic regression analyses showed that a small set of static measures involving word identification fluency, phonological awareness, rapid digit naming, and oral vocabulary differentiated students who became RD from typical readers with a sensitivity of 78% and a specificity of 79%, with an AUC of .86. It should be noted, however, that the

reported sensitivity is probably an overestimation of the true rate. Because only low-functioning students were included in the sample, classification accuracy did not consider those higher functioning classmates who may have passed the screen but developed RD. Additionally, the combination of measures found to be most predictive would take a prohibitive amount of time to be used as a screening measure.

Investigating More Efficient Screenings

Although multivariate screenings have proven to be accurate, they have not been identified as efficient in an RTI framework. With the increased implementation of RTI in the United States (Berkeley, Bender, Peaster, & Saunders, 2009), educational units have moved toward the use of single measures or a more limited combination of measures in their universal screening process.

One such screening measure is the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kiminski, 2003). DIBELS is a set of timed tasks appropriate for students in kindergarten through sixth grade that were designed to measure the acquisition of literacy skills. (The term “dynamic” refers to the fact that the assessment has multiple forms that can be administered repeatedly across grades, and thus, differs from the way the term will be later defined in this research study.) The DIBELS was designed as a progress monitoring tool that would measure growth on a frequent and ongoing basis. Currently, DIBELS is more commonly used as a universal screening measure to identify students as at risk for RD (US Department of Education, Office of Inspector General, 2007). Two of the most widely used subtests of the DIBELS are Initial Sound Fluency (ISF) and Nonword Reading Fluency

(NWF). The former is a measure of phonological awareness, and the latter is a measure of decoding ability. Recent research found that the use of DIBELS as a screening measure resulted in a high number of students incorrectly identified as at risk (e.g., Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009). Catts et al. (2009) examined longitudinal data from over 17,000 students in Reading First schools from kindergarten to third grade. The data included five measures from the DIBELS as predictor variables in addition to two reading achievement outcome measures. Results showed that DIBELS measures were characterized by floor effects and poor predictability during each of the measure's initial administrations. The presence of high false positive rates most likely resulted in schools using resources on students who did not require extra assistance and would have instead learned in the regular classroom environment.

Foorman et al. (1998) also attempted to increase the efficiency of early identification with the development of the *Texas Primary Reading Inventor (TPRI)*. They identified 945 students in kindergarten, first, and second grades and administered measures of reading and reading-related skills four times a year for one to three years. They reported that a brief measure of a combination of letter-sound recognition and phonological awareness, given in December of kindergarten, was the best predictor of end of first-grade reading outcome (word reading and reading comprehension). This screening had a sensitivity of 91%, but a specificity of only 63%. Comparable levels of identification accuracy were found when using measures of word reading and phonological awareness, administered in October of first grade,

to predict end of first grade reading achievement (indexed by a composite of word reading and reading comprehension; sensitivity 93%, specificity 63%). Although this screening measure was highly sensitive, it is important to note that the screening was not administered until December of kindergarten; it is likely that the number of false positives would have increased if the screening was administered in the beginning of kindergarten.

Dynamic Assessment

The above screenings utilized traditional *static* measures; the term refers to assessments of already learned products or abilities at one point in time (Lidz, 1991). In a static assessment, students answer a set of items with little or no feedback. Because of the high error rates associated with current static screenings, educators and researchers have proposed *dynamic assessment* as an alternative assessment or screening method. Dynamic assessment refers to a variety of procedures that embed interaction with a child as part of the assessment process. In dynamic assessment, the examiner takes an active role by teaching a task or providing explicit prompts. Success is measured by both a student's level of independent performance as well as a student's assisted performance (i.e., progress). Dynamic assessment takes into account both the process and the product of learning; in other words, it considers growth in response to some sort of instruction (for a review, see Sternberg & Grigorenko, 1998). Supporters of dynamic assessment believe that it can provide information about a child's ability to respond to instruction that is not obtainable through more traditional assessment sources (e.g., Feuerstein, Haywood, Rand,

Hoffman, & Jensen, 1982; Grigorenko & Sternberg, 1998; Lidz, 1991). Recently, Grigorenko (2009) summarized three assumptions shared among researchers interested in the usefulness of dynamic assessment for educational purposes: (1) static assessment might not adequately capture the wide range of educational experiences that young children bring to formal schooling; (2) educators should be most interested in the potential growth a student can make, not where the student is at the time of assessment; and (3) assessment should provide information related to selecting or modifying appropriate instruction. Lidz (2005) reviewed the current status of dynamic assessment and noted that the approach addressed the idea of responsiveness to intervention. In the review, Lidz suggested that the student's response to intervention embedded within the assessment procedure provides evidence for instructional planning.

The theoretical roots of dynamic assessment are based on the work of Lev Vygotsky and his idea of the *zone of proximal development* (Vygotsky, 1978). The zone of proximal development has been defined as the difference between a child's performances on an unaided task compared to his or her performances on a task when guided by a more experienced adult. This difference can be related to the notion of readiness, in that a child with a wide zone may be more likely to achieve success on tasks when given some guidance. According to Vygotsky, skills in the child's zone have not yet fully emerged, but can emerge with adult feedback/instruction. Vygotsky did not intend for an approach utilizing this notion to replace more traditional measures; instead, he proposed its use as a supplement to static

assessment. Later, Reuven Feuerstein proposed dynamic assessment as a means to measure a child's ability to profit from instruction, an experience he termed *mediated learning experience* (Feuerstein, 1979; Feuerstein & Rand, 1974; Feuerstein, Rand, & Rynders, 1988). Researchers interested in dynamic assessment have suggested that there are substantial numbers of children whom, due to factors such as cultural differences or lack of a mainstream academic experiences, have their actual capabilities underestimated by static assessments.

Over the years, many different dynamic assessment methods have been suggested. Some of the most common methods of dynamic assessment include Feuerstein's *Learning Potential Assessment Device* (LPAD), the test-teach-retest, and graduated prompts. The first has a strong clinical orientation, while the latter two place an emphasis on predicting achievement and, more recently, educational placement. However, an important factor in all methods is the emphasis on examiner instruction and the subsequent change in the student's performance.

The work of Feuerstein (Feuerstein, 1979; Feuerstein & Rand, 1974; Feuerstein, Rand, & Rynders, 1988) attempted to address criticism of static assessments and their potential to measure the ability of students from different cultural backgrounds. The development of the LPAD resulted from this criticism (Feuerstein, Rand, & Hoffman, 1979). This measure was designed to help a child self-modify cognitive processes with assistance from a more-experienced adult. The subtests in the LPAD measure broad cognitive skills like reasoning and memory strategies. This is consistent with most traditional dynamic assessment measures in

that general cognitive ability, rather than specific content areas, are assessed.

Although this is an assessment, the focus is not on the prediction or diagnosis of a disability. Instead, the examiner focuses on the remediation of the child. Examiners administering this assessment provide contingent prompts based on each child's individual performance. Because the instruction is individualized to each child's specific needs, one criticism of this method is that this assessment requires extensive examiner skill and training. Without standardized procedures, the technical characteristics are not easily assessed.

The test-teach-retest assessment was developed primarily by Budoff (1974, 1987) and is also known as *learning potential testing*. In this type of dynamic assessment, examiners use a test to identify areas of deficit or emerging skills that are possibly related to lack of experience. An intervention is then provided to children that targets the area of need. Finally, a retest is administered, with this performance serving as a measure of the children's modifiability. This type of assessment was first proposed by Budoff (1974) as a way to address assessment issues related to children's varied past experiences. He felt that the instruction provided by the examiner was a means to equalize these differences in experience that could adversely affect a child's performance on a standardized test. More recently, it has been adopted by researchers as a way to assess children with disabilities, children from non-mainstream cultural backgrounds, and children with low socioeconomic status (Feuerstein, 1979; Tzuriel & Klein, 1987; Pena, Gillam, Malek, Ruiz-Felter, Resendiz, Fiestas, et al., 2006). Budoff's research included fairly standardized training and instruction procedures,

whereas others advocated for a more unstructured method (e.g., Tzuriel & Klein, 1987; Gutierrez-Clellen, Pena, & Quinn, 1995).

In a *graduated prompts* approach, examiners provide children a series of progressively explicit prompts until the child is able to solve the task. These prompts are typically standardized and are administered in a set order depending on the child's response. This approach was used by Campione (1989) and Campione and Brown (1987). The graduated prompts approach typically relies on new content rather than the type of complex tasks that are often seen on traditional tests. Introducing new content is thought to aid the standardization of prompts, which should lead to increased reliability, and also allow for greater differentiation of children in the lower end of the distribution (Grigorenko & Sternberg, 1998). Consequently, one of the primary interests of the developers of a graduated prompts approach is its ability to predict later achievement levels (Campione & Brown, 1987). Some evidence has found that the number of prompts needed to successfully achieve a task contributed to variance in performance of a task (Campione & Brown, 1987). However, more research is needed on its utility in predicting academic achievement at a later date. One possible limitation of this approach is that the feedback provided is predetermined, and therefore, not modified to meet an individual child's unique learning needs. However, the graduated prompts approach has been shown to produce useful information regarding children's ability to benefit from instruction and to transfer that learning within the task domain in which the learning occurred (Missiuna & Samuels, 1988).

Curriculum-Based Dynamic Assessment

A more recent development in the history of dynamic assessment is the development of assessments that are domain- or curriculum- specific. In fact, inclusion of actual curriculum content into a dynamic assessment is clearly a deviation from a more traditional approach. However, advocates of curriculum-based dynamic assessments insist that if an assessment is to be relevant to an academic setting, then the content must be directly linked and applicable to educational content (e.g., Haywood & Lidz, 2007).

This curriculum-based notion of dynamic assessment can be directly related to academic settings and seems to lend itself particularly well to the screening of students to determine risk status. Results on a dynamic assessment can be used as an indicator of how well a child might perform when given instruction in the classroom. Children come to school with various levels of exposure to literacy, and one reason children fail an early literacy screening may be based on their limited early literacy exposure. When these same children are provided with classroom instruction, they respond well and therefore are not thought to be truly at risk for RD. Additionally, some children may perform poorly on a test due to difficulty comprehending directions. The models and/or prompts provided by the examiner in a dynamic assessment may benefit such children. By measuring how well students respond to feedback during dynamic assessment, examiners may be able to gauge how easily a child would learn to read given a longer period of instruction in the classroom.

As noted by Fuchs, Compton, Fuchs, Hollenbeck, Craddock and Haddock (2008), dynamic assessment might effectively discriminate students performing at the lower end of the distribution. For example, Fuchs and colleagues suggested that two students with identical low scores on a screening of risk for RD might not actually have the same potential to develop RD. Instead, it is possible that the two students differ in the amount of assistance/modeling required to learn a task. The student who struggles initially, but then is successful when provided with assistance, might have a high potential to learn within the classroom. The student who does not learn a task even with a high level of assistance might not learn to read without extensive, individualized intervention. This dovetails nicely with Grigorenko's (2009) observation that in an educational setting, teachers should be more interested in a student's potential for success following quality education than in a student's ability to succeed during a particular testing period.

Dynamic Assessment as a Screening for Reading Disabilities

One of the first researchers to utilize dynamic assessment to predict reading achievement in young children was Spector (1992). She hypothesized that a dynamic assessment of phonological awareness would predict later reading achievement with greater accuracy than similar static assessments. In this study, 38 kindergarten students were administered the following static assessments in the fall of kindergarten: phoneme segmentation, phoneme deletion, invented spelling, and a receptive vocabulary assessment. In addition, a dynamic assessment of phoneme segmentation was administered. This task was similar to the static phoneme

segmentation task except the examiner provided feedback and increasingly supportive prompts after each missed item. The fixed graduated prompts included the examiner a) pronouncing the target word slowly, b) asking the child to say the first sound in the word, c) telling the child the number of sounds in the word, d) modeling segmentation using tangible objects, and e) providing hand-over-hand assistance to the child while pronouncing the phoneme segments. Items were given a point value based on the number of prompts needed for the child to answer the item correctly, with six being the highest score obtained per item. This dynamic assessment is consistent with a graduated prompts framework, with the measure indicating the degree of independence the child achieved during the assessment.

Significant moderate correlations were noted between all of the phonological awareness measures (phoneme segmentation, phoneme deletion, invented spelling and dynamic phoneme segmentation) administered in the fall. The correlations of the dynamic phoneme segmentation with each of the three remaining static measures were greater than the correlations of static measures. Additionally, the dynamic phoneme segmentation correlated more highly with the spring reading performance than any of the static predictors. Multiple regression analyses were completed to determine which fall measure best predicted spring phonological awareness and word recognition scores. Results indicated that the dynamic measure contributed between 12% and 14% of the unique variance on a phonological awareness measure and 21% of the unique variance on a word reading measure; the dynamic measure was the only significant predictor of word reading.

As previously discussed, O'Connor and Jenkins (1999) screening battery, administered in October of first grade, predicted end of year reading with a sensitivity of 100% and a specificity of 87%. In an attempt to improve their prediction accuracy, a dynamic segmentation task was administered to students in the third cohort who scored below 80% on the corresponding static measure in addition to the first grade static measures. This dynamic task included a series of standardized prompts to teach students to segment words into onsets and rimes. The prompts included a) an examiner modeling the task while using Elkonin boxes (Elkonin, 1973), b) the child segmenting using the Elkonin boxes without a teacher model, and c) an examiner administering a trial without any prompts or Elkonin boxes. Both the total number of prompts needed by a child as well as the total words correctly segmented was measured. When the results of this assessment were combined with those from the static assessments, specificity improved from 87% to 96%. The authors concluded that this dynamic application of a segmentation task reduced the floor effects which led to this reduced over-selection rate. However, the dynamic assessment did add a substantial time element (approximately 30-35 minutes) to the assessment procedure, which would prohibit it from being administered to a large number of students as a universal screening measure.

A recent study (Caffrey, 2006) examined the predictive validity of dynamic assessment as compared to progress monitoring, which has also been utilized as a measure of growth predicting later RD (Compton et al., 2006). In this study, 120 students in kindergarten (N=25) and first grade (N=95) were administered a dynamic

screening of nonword reading in December, and progress monitoring measures were collected starting in mid-January. For the kindergarten students, the progress monitoring assessment was based on letter identification, and for first-grade students the progress monitoring assessment was word reading. The outcome measures collected in April and May of the same year consisted of word reading, nonword reading, fluency, and spelling measures.

This study utilized a dynamic assessment measure of word reading consisting of three subtests that required a child to learn a decoding rule, such as short versus long vowels. In this assessment, if a child did not master the content of the first subtest by reading 5 of 6 words correctly, he or she moved to level 2. At this level, an examiner prompt was provided that was directly related to the decoding rule. If the child still did not master the content, increasingly explicit prompts were provided until either mastery was reached or all prompts had been given (Level 5). The next subtest was only administered if the child reached mastery at some point. The subtests were scored 1 through 5 points, with 1 indicating mastery at the first opportunity (i.e., without prompts) and 5 indicating a need for prompts up to the last level. Therefore, the best score on the assessment was three and the poorest score was 15.

A commonality analysis was conducted, which allowed the author to determine the unique contribution of the following predictors administered in December: (1) dynamic screening of nonword reading, (2) progress monitoring score at the first time point (intercept), and (3) progress monitoring slope. Results for the kindergarten sample showed that, although the intercept explained the most unique

variance in each of the dependent measures, fall dynamic assessment contributed a significant, yet small, amount of unique variance to word identification and a large amount of unique variance to word attack. In the first grade sample, fall dynamic assessment explained a significant amount of unique variance in each of the dependent measures.

Limitations of Past Research on Dynamic Assessment

One criticism of dynamic assessment is a lack of evidence related to its validity. However, results from these studies provide initial evidence of its predictive validity for later RD and indicate that dynamic assessment might be a viable supplement to static assessments. None of those using dynamic screening advocate its use as the sole assessment/screening measure. Instead, dynamic assessment may be most useful as a supplemental measure to a static screening. This could be completed in a two-step process. First, an efficient static screening measure could be administered to all students, with a cutoff score chosen to ensure high sensitivity. Next, a follow-up dynamic screening measure would be administered to those students who failed the screening. The purpose of administering the dynamic screening would be to reduce the number of false positives. It is possible that dynamic assessment might be most beneficial in differentiating low-achieving students. Static test items are typically scored as either right or wrong, whereas dynamic assessment provides more than one opportunity to answer an item correctly. This might be particularly useful in an early literacy skill such as phonological

awareness, which is affected by early literacy exposure; low-achieving students might be differentiated based on how easily they learn the skill.

Although the above research provides support for the use of dynamic assessment in predicting later reading ability, further research is needed to examine its usefulness as part of a screening protocol. First, a dynamic assessment used for screening purposes should be able to be administered efficiently (i.e., in a limited amount of time). Although there are no “standard guidelines” for the length of time a screening measure should require, a review of state education department guidelines indicated that ten minutes or less was a common guideline. This is particularly important if a dynamic assessment is to be used as a supplemental screening measure to follow a static measure. Secondly, it is imperative that the reliability of administration be addressed in future studies. For a dynamic assessment to be utilized as a universal screening measure, it must be given in exactly the same manner to all students. In the studies reviewed above, the dynamic assessments were administered by researchers with substantial experience with reading development and assessments. However, this is probably not true of personnel administering screenings in a school setting. Therefore, a screening should be easily administered, scored, and interpreted by individuals, regardless of their experience administering assessments. As stated by Jenkins (2003), a screening measure should be “simple enough to be implemented on a wide scale, by normal people under normal circumstances,” and should not require a specialist (e.g., school psychologist, speech-language pathologist, certified reading specialist) for administration and interpretation. Finally, a dynamic

screening measure that can be given efficiently and reliably must also provide information related to the potential risk of a student to develop RD at a later time.

Overview of the Present Study

The purpose of this research was to investigate the use of a dynamic screening of phonological awareness in the early identification of students at risk for RD, with emphasis on investigating the value of adding dynamic assessment over and above traditional assessment. Study I examined the fidelity and reliability of the dynamic screening when administered by school personnel to 372 students at the beginning of kindergarten. For the dynamic assessment to be utilized as a standardized screening measure, it was important that high levels of reliability, examiner fidelity, and validity were obtained. Additionally, the distributional characteristics of the assessment were examined and compared to those of static phonological assessments. Of specific interest was the presence or absence of indications of a positively skewed distribution in the screening measures, which is often indicative of a floor effect.

The purpose of Study II was to investigate the use of a dynamic screening of phonological awareness to identify students at risk for RD. Based on results of Study I, a revised version of the dynamic screening was developed, and the revised version was administered to two samples of kindergarten students (N=90 and N=161) in September of kindergarten. These data were obtained to ascertain in part if the dynamic screening demonstrated good distributional characteristics. A second goal of this study was to investigate the value added of dynamic assessment over and above traditional static measures. Logistic regression analyses examined the predictive

validity of the revised dynamic assessment, by itself and in combination with static screening measures, in predicting end of the kindergarten reading outcomes. Of particular interest was investigating the classification accuracy of the dynamic assessment as a supplemental measure to a commonly-utilized static measure.

In summary, the current studies will address the following research questions:

1. Can a dynamic screening measure of phonological awareness be used with high reliability and fidelity by school personnel?
2. How do the distributional characteristics of a dynamic screening measure of phonological awareness compare to those of static measures of phonological awareness when administered at the beginning of kindergarten?
3. Does a dynamic screening measure of phonological awareness administered in kindergarten add significantly to the prediction of reading risk over and above a comparable static measure of phonological awareness *utilizing the same items*?
4. Does a dynamic screening measure of phonological awareness administered in kindergarten add significantly to the prediction of reading risk over and above a *commonly used* static measure of phonological awareness?

CHAPTER II: STUDY I

Study I addressed the following two research questions: *Can a dynamic screening measure of phonological awareness be used with high reliability and fidelity by school personnel? How do the distributional characteristics of a dynamic screening measure of phonological awareness compare to those of static measures of phonological awareness when administered at the beginning of kindergarten?*

To address the first question, observations were completed by the PI, and inter-rater reliability, test-retest reliability, and procedural fidelity estimates were calculated. To address the second question, the distributional characteristics of the dynamic screening were compared to those of other phonological awareness screening measures administered during the same time period.

Method

Participants

Participants in this study were 372 kindergarten students recruited from the Lawrence public school system. This school district was somewhat diverse in terms of ethnicity (approximately 63% Caucasian, 11% African-American, 6% Hispanic, 7% American Indian/Alaskan native, 6% Asian/Pacific Islander, and 7% multi-racial). Thirty-five percent received free or reduced lunch. The majority of the students (N=237) were selected on the basis of risk status for RD. Any child classified as “Some risk” or “At risk” (i.e., a score below 8) based on the Initial Sound Fluency subtest of the Dynamic Indicators of Basic Early Literacy (DIBELS; Good & Kaminski, 2002) was recruited. A second group of 135 kindergarten students

were randomly selected from those students found not to be at risk for RD based on the Initial Sound Fluency measure. Exclusionary criteria included the following: a designation of “nonverbal” on an Individualized Education Plan; limited English proficiency as indexed by a score of 1 or 2 on the oral language subtest of the PreLAS (Duncan & DeAvila, 1998); or significant health or cognitive impairment (e.g., mental retardation, hearing impaired, autism). Signed informed parental consent statements were not required because the assessment was adopted as part of a school-wide early identification procedure.

Measures

Three phonological measures were included in this study. The Initial Sound Fluency subtest of the DIBELS was administered. Two additional research-generated phonological assessments were administered. These assessments consisted of the same items, but one was given in a static manner and the other was administered in a dynamic manner (see below).

Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Initial Sound Fluency (Good & Kiminski, 2003). The *Initial Sound Fluency (ISF)* subtest of the DIBELS is a measure of a student's ability to recognize and produce the initial sound in an orally presented word. The examiner showed the student four pictures, named each picture, and asked students to identify (i.e., either point or say) the picture that began with the sound produced by the examiner. Students were also asked to produce the beginning sounds of words presented orally by the examiner. The amount of time taken to identify/produce the correct sounds was converted into the number of initial

sounds correct in a minute. Alternate-form reliability of this measure is .72 (Good et al., 2004) and test-retest reliability is .66. (Catts et al., 2009).

Static Deletion Task (SDT). This task is similar to one used clinically (Catts, 1999). A deletion task was chosen because research has shown it to be one of the best phonological awareness tasks in the prediction of later reading achievement (Gillon, 2004; Kroese, Hynd, Knight, Hiemenz, & Hall, 2000; Muter, Hulme, Snowling, & Stevenson, 2004; Schatschneider, Francis, Foorman, Fletcher, & Mehta, 1999). For this task, students were asked to say a word produced by the experimenter and then repeat the word after deleting a syllable or a phoneme specified by the examiner. The correct response always formed a real word after the deletion was performed (i.e., “Say cowboy without cow.”) The SDT consisted of four sets of words, with each set increasing in complexity. In the first set, the first syllable from a two-syllable word (either compound or a word with a prefix) was deleted. In the second set, a syllable from a non-compound, two-syllable word was deleted. In the third set, the initial consonant from a single-syllable CVC word was deleted. In the fourth set, the first consonant from a single-syllable CCVC word was deleted. Administration of this assessment was discontinued when students received a score of 0 on five consecutive items. There were 16 total points possible.

Each item was a real word, and the target response, after the initial phoneme was deleted, was also a real word. The rationale for selecting initial phoneme deletion was that research has shown initial sounds to be easier to manipulate than non-initial sounds (Chafouleas, VanAuken, & Dunham, 2001; McBride-Chang, 1995; Stahl &

Murray, 1994). Additionally, the first few items that required phoneme deletion had a continuant in the initial position of the word (e.g., /s/), which researchers have suggested is easier than phoneme deletion with a stop (e.g., /p/) in the initial position (Ball, 1993; Lewkowicz, 1980; Marsh & Mineo, 1977). See Appendix A for the list of items. The deletion task took approximately 5 minutes to administer.

Dynamic Screening of Phonological Awareness (DPSA). This assessment included the same items included in the SDT described above, but it was administered in a dynamic manner, with graduated prompts provided to make the task more appropriate for kindergarten students. The dynamic measure always followed the static measure in terms of the order of presentation of tasks. In addition, to reduce redundancy the examiner started the DSPA on the first item that the child had missed on the SDT. Items prior to the starting item were given the full credit. For example, if a child missed item number 2 on the static version, then the examiner started with item number 2 on the DSPA. This was true even if the child answered any one of the items 3-16 correctly after missing item 2 on the static assessment. This procedure was used so that the two assessments could be directly compared (i.e., the same items were used for both measures).

The protocol for the prompts on the DSPA was the following. According to the assessment procedures, when students gave a correct response, the response was acknowledged as such (i.e., “that’s right”). Alternatively, when students gave an incorrect response to an item in each set, the examiner provided a series of prompts until the item was answered correctly or until the answer was provided by the

examiner. See Appendix B for the prompt protocol. Each item was assigned 0 to 4 points, with a higher score on an item indicating the need for fewer examiner prompts. Therefore, the procedure for scoring reflected the degree of independence that a student achieved in performing the task. There were 16 items; the highest score possible was 64 points. Administration of the DSPA was discontinued if the child received a score of 0 on five consecutive items. Although administration time varied due to the number of prompts provided to an individual student, the average time of administration was 8-10 minutes.

Procedures

Examiner training. All examiners participated in a two-hour training session conducted by the author and her research mentor. First, the author and mentor provided a brief overview of the SDT and the DSPA. Following this overview, administration of the two measures was modeled. Next, the school personnel practiced administering both measures; during this time, the author and mentor observed the administrations and provided feedback and clarification when appropriate. In the last portion of the training session, a brief review was provided by the author. The author contacted the examiners the day after training via email to reiterate important features of the assessment and to answer any questions posed by the examiners.

The Lawrence public school district required all school personnel who administered the DIBELS measures to participate in a three-hour training session. This training was conducted by two reading specialists employed by the school

district. The training format was similar to that of the DSPA.

Administration schedule. In this study, all testing was completed by school personnel, which included Title 1 certified teachers, reading specialists (i.e., degree in reading), or individuals without a four-year college degree who had participated in district training related to reading assessment and instruction. The ISF, the SDT, and the DSPA were administered to the students within the first 3 weeks of kindergarten. All measures were individually administered in the students' regular school environment.

Data Scoring and Data Entry. Data scoring and entry of the ISF was completed by school district personnel per their standard protocol. The SDT and the DSPA were originally scored by the school district examiner who administered them. The author double checked the scoring of the latter two assessments. If an error in addition was made by the original examiner, the correct score was written on the form, and this value was used in data analyses.

Results and Discussion

All data analyses were completed using SPSS 16.0 (2007). All analyses were based on raw scores.

Reliability and Fidelity

First, the following research question was addressed: *Can a dynamic screening measure of phonological awareness be used with high reliability and fidelity?* To answer this question, inter-rater and test-retest reliability, as well as procedural fidelity, were obtained.

Reliability is expressed as a statistical index with values ranging from 0 (not at all reliable) to 1 (perfectly reliable). In applied settings where important decisions are based on a specific test score (e.g., a school setting), a reliability of .90 is considered “minimally acceptable” and a reliability of .95 is the “desired” standard (Nunnally, 1978). Salvia and Ysseldyke (1988), however, state that a level of .80 is acceptable for the purpose of a universal screening that will lead to some sort of further assessment or diagnostic process.

Procedural reliability, or fidelity, reflects the extent to which examiners adhere to the procedural requirements of a given assessment. For this study, a fidelity protocol was developed by the PI to measure the accuracy of the examiners’ administration of the DSPA. The fidelity checklist included 12 items (see Appendix C). The author observed 43 administrations, which equaled 12% of the DSPA administrations completed. Fidelity checklists were scored online by the PI at each administration. At least two different examiners were observed at each school, but it was not possible to observe every examiner due to time and scheduling constraints. Items on the fidelity checklist were scored as correct or incorrect based on the observation. Fidelity was calculated as the ratio of total points obtained to total points possible. Results indicated that trained school personnel gave the task with high fidelity (90%).

In addition to the fidelity checklist, scoring reliability was also obtained. Scoring reliability involved the extent to which examiners followed the established scoring procedures. At each observed DSPA administration, the PI scored the DSPA

online, and the score was compared to the score obtained by the original examiner. Inter-rater reliability was calculated as the percent of agreement per item (scored as either correct or incorrect) between the raters. Results indicated high inter-rater reliability (.98).

Finally, test-retest reliability of the DSPA was examined. Test-retest reliability was estimated by administering the DSPA to the same students on two different occasions. Fifty two students (14%) were randomly selected after the initial administration, and examiners re-administered the DSPA to those students three weeks later. The students' performance at the first administration was correlated with their performance at the second administration; the results showed good reliability (.86). One concern is with the possible artificial inflation of the reliability estimates based on the floor effect seen in the DSPA. Therefore, test-retest reliability was also calculated on those students in the original 52 who did not score a zero on the original administration (N=26); this test-retest reliability was .69. This level of reliability is to be expected when examining test-retest reliability of a dynamic assessment. As Lidz (1991) discussed, an assessment that is highly stable from one administration to another is, by definition, not dynamic.

Taken together, these reliability estimates indicate that the DSPA can be administered by school personnel in a uniform manner. This is particularly noteworthy as the examiners had little to no experience with dynamic assessment prior to this research study, and although a training session was provided, it was limited in scope. Therefore, the standardized approach of this measure seemed to

make the DSPA easy for the examiners to administer, regardless of educational level or experience with administering assessments.

Distributional Characteristics of the DSPA

The following results addressed the second research question: *How do the distributional characteristics of a dynamic screening measure of phonological awareness compare to those of static measures of phonological awareness when administered at the beginning of kindergarten?* In order to make comparisons between the DSPA, the SDT, and the ISF, the distributional characteristics of the measures were first examined. Descriptive statistics are presented in Table 2.1; all scores presented are raw scores.

Table 2.1
Descriptive statistics for predictor variables

	Mean	SD	Median	Range	Skewness	Kurtosis
DSPA	23.25	21.20	24.0	0-64	0.25	-1.19
SDT	3.28	4.20	1.0	0-15	1.22	0.86
ISF	7.53	6.56	7.0	0-27	1.58	2.60

Note. DSPA=Dynamic Screening of Phonological Awareness; SDT= Static Deletion Task; ISF= Initial Sound Fluency

Recall that one concern regarding screening measures of phonological awareness is the presence of a floor effect. The skewness statistic, an index of the symmetry of the distribution, can provide information regarding the presence of floor effects in the data. The farther the absolute value of the skewness statistic is from zero, the greater the skew of the distribution. Positive numbers, typically seen in a distribution with a floor effect, are indicative of a mean that is closer to the lower end of the distribution

and an asymmetric tail extending toward the higher end of the scale. The skewness values for the DSPA, SDT, and the DIBELS ISF were .247, 1.22, and 1.58, respectively. See Figure 2.1 for histograms of the three variables. Further inspection of the data revealed that 205 students scored a zero on the SDT. Of those students, 68 scored 1 SD *above* the mean on the DSPA. These results suggest that the feedback/learning associated with the dynamic screening measure assisted some students in learning the deletion task. However, many students who scored a zero on the SDT (147) still performed less than 1 SD below the mean on the DSPA.

In a dynamic assessment, the goal is to provide an appropriate amount of scaffolding or instruction so that a student who performs poorly at the onset can demonstrate at least some improvement over time. Whereas the DSPA appeared to assist many students in achieving better scores on the phonological awareness task, there were still a number of students with scores at the low end of the distribution. However, the skewness statistic associated with the DSPA indicated less of a floor effect than those of the SDT and the ISF, suggesting its promise as a potential screening measure for risk for RD. Thus, the DSPA was revised; distributional characteristics and predictive validity of this revised version are discussed in Study II. Therefore, the predictive validity of the DSPA was not addressed in this study; instead, additional data were collected at a later time on a revised version of the DSPA (see Study II).

CHAPTER III: STUDY II

The dynamic screening measure was revised based on the results of Study I. Specifically, additional examiner models and prompts were included and the total number of items was changed to twenty. These revisions were made in an attempt to reduce the number of students scoring at the low end of the distribution on the dynamic screening measure. After revisions were made, but prior to the start of Study II, pilot data were collected to assess the distributional characteristics of this revised measure. The Dynamic Screening of Phonological Awareness-Revised (DSPA-R; Bridges & Catts, 2008) was administered to a sample of students (N=102) in January of kindergarten; the sample came from two Midwestern school districts. An inspection of the data showed that the distribution of scores associated with the DSPA-R approximated a normal distribution, with a skewness statistic of -.08. However, since the dynamic screening measure was administered four months later than in Study 1, it was not clear whether the revision or the time of administration improved the distributional characteristics.

Study II addressed the following research questions: *How do the distributional characteristics of a dynamic screening measure of phonological awareness compare to those of static measures of phonological awareness when administered at the beginning of kindergarten? Does a dynamic screening measure of phonological awareness administered in kindergarten add significantly to the prediction of reading achievement over and above a comparable static measure of phonological awareness utilizing the same items? Does a dynamic screening measure of phonological*

awareness administered in kindergarten add significantly to the prediction of reading achievement over and above a commonly used static measure of phonological awareness? To address the first of these questions, the distributional characteristics of the DSPA-R were compared to those of the SDT and the ISF administered during the same time period. To address the second and third questions, regression analyses were conducted to investigate the ability of the DSPA-R to add significantly to the prediction of end of kindergarten reading outcomes over the SDT as well as the ISF. Additionally, the predictive ability of the DSPA-R as a supplemental screening measure to the ISF was examined.

Method

Participants

Participants in this study were 251 kindergarten students from four school districts in the Midwest. Because the districts were located in two different geographic regions, the large sample was divided into two smaller samples. Sample 1 (N=90) represented students from three small school districts in Iowa. This population included approximately 95% Caucasian students, with between 15-20% of the students in the three schools receiving free or reduced lunch. Information provided from the school districts indicated that thirteen (14%) students were deemed to be at risk. This was based on performance on the ISF administered at the beginning of kindergarten.

Sample 2 (N=161) was comprised of students from a large school district in Lawrence, Kansas; demographic statistics for this district were provided in Study I.

All participants in Sample 2 were part of a larger University of Kansas research study investigating early identification of RD. School records indicated that 112 were deemed to be “at risk” for RD based on performance on the ISF and Letter Name Fluency subtests of the DIBELS. The analyses utilizing Sample 2 data were conducted using both the full sample of 161 students (labeled “the full sample”) as well as a reduced sample of 96 students (labeled “reduced sample”). The reduced sample did not include sixty-five students who were randomly selected to receive intensive intervention provided by members of the University of Kansas research study through the kindergarten year. The latter children were excluded in one set of analyses because it is yet to be determined if the intervention provided to these students yielded significantly different outcomes than those from the students at risk who received school-administered intervention.

Exclusionary criteria in both samples included the following: a designation of “nonverbal” on an Individualized Education Plan (IEP); limited English proficiency as indexed by a score of 1 or 2 on the oral language subtest of the PreLAS (Duncan & DeAvila, 1998); or significant health or cognitive impairment (e.g., intellectual disability, hearing impaired, autism). In addition, a limited number of additional students from Sample 1 were excluded by the teachers due to one of the following: (1) limited English proficiency levels per teacher report, (2) behavior concerns, or (3) an IEP designating a cognitive delay or severe speech impairment. Because the assessment was adopted as part of school-wide early identification procedures in all

participating schools, a signed informed consent from the parents of the participants was not required.

Measures

The following measures were utilized as predictor variables in this study: the ISF, the SDT (Sample 1 only), and the DSPA-R. Additionally, the Word Identification and the Word Attack subtests from the Woodcock Reading Mastery Tests-Revised/NU (WRMT-R/NU; Woodcock, 1998) were administered to all students; these measures were used in all analyses as outcome variables.

DIBELS ISF. This task was described in Study I.

Static Deletion Task (SDT). This task was described in Study I. The only difference was the inclusion of four new items that were also included in the DSPA-R (see below); therefore, there were 20 points possible. Only students in Sample 1 were administered this measure. This task was not part of the assessment battery administered to students in Sample 2 as part of the larger research study.

Dynamic Screening of Phonological Awareness (DSPA-R; Bridges & Catts, 2008). This measure was almost identical to the measure used in Study I. However, this version included four more items than the previous version. Items were added to the beginning and the end of the measure in order to increase the possible range of scores. See Appendix D for the list of items for the DSPA-R. Two additional changes were made to this assessment, with the primary goal of reducing the floor effect seen in the DSPA. First, the DSPA-R utilized an additional training item at the beginning of the test, in which students had the opportunity to provide an answer and receives

standardized feedback related to their response. The second change was related to prompt 3 for items 1-6. For the DSPA, the visual component of prompt 3 was two black squares, which the examiner used to “tap out” syllables or phonemes of a word. In the DSPA-R, for items 1-4, there was a colored drawing in each square to represent each syllable of the compound word. For example, for item 1 (doghouse), there was a colored drawing of a dog in the left square and of a house in the right square. For items 5-6, the square on the left was black to represent the prefix, and the square on the left contained a colored drawing of the remaining syllable (i.e., night for item 5). These changes were made in order to provide a higher level of scaffolding as students learned the task.

Woodcock Reading Mastery Test-Revised/NU Word Identification

(Woodcock, 1998). The Word Identification subtest was a measure of untimed real word reading in isolation. Students were required to read a list of words that gradually increase in length while at the same time decreasing in frequency of occurrence. The WRMT-R/NU manual reports the split-half reliability is .98.

Woodcock Reading Mastery Test-Revised/NU Word Attack (Woodcock,

1998). The Word Attack subtest was a measure of untimed nonsense-word reading in isolation. This measure assessed a student’s ability to apply grapheme-phoneme correspondence rules in order to pronounce unfamiliar printed words (i.e., pronounceable nonwords varying in complexity). Students were required to read a list of increasingly complex nonwords. The manual reports the split-half reliability is .94.

Procedures

Administration schedule. For Sample 1, school district personnel administered the ISF, and the PI of this study administered the SDT, DSPA-R, Word Identification, and Word Attack measures. For Sample 2, all measures were administered by personnel affiliated with the larger research project. All measures were individually administered in the students' home schools. Table 3.1 displays each measure administered and the corresponding time of administration for Sample 1 and Sample 2.

Table 3.1
Administration schedule for Sample 1 and Sample 2

Measure	Time administered: Sample 1	Time administered: Sample 2
ISF	September 23-30	September 8-12
SDT	September 8-12	NA
DSPA-R	September 8-12	September 15-October 3
Word Identification	April 27-April 30	April 27-May 8
Word Attack	April 27-April 30	April 27-May 8

Note. ISF= Initial Sound Fluency; SDT= Static Deletion Task; DSPA-R= Dynamic Screening of Phonological Awareness-Revised

Results

Analyses were conducted using SPSS 16.0 (2007). For all analyses, raw scores were utilized.

Distributional Characteristics of the DSPA-R

This section addressed the following research question: *How do the distributional characteristics of a dynamic screening measure of phonological awareness compare to those of static measures of phonological awareness when administered at the beginning of kindergarten?* First, descriptive statistics for the

predictor variables used in this research study are presented in Table 3.2. Again, the statistic of primary concern was the skewness of the distribution. The DSPA-R and the ISF had skewness statistics close to zero (i.e., -.062 and .187). The skewness statistic associated with the SDT measure (.990) showed that the data were skewed positively. These results suggested a floor effect in the SDT that was not noted in the ISF or DSPA-R. See Figure 3.1 for histograms of the predictor variables.

Table 3.2
Descriptive statistics for Sample 1 (N=90)

Measure	Mean	SD	Median	Range	Skew	Kurtosis
ISF	16.86	8.81	16.0	0-40	0.187	-.312
SDT	4.73	5.43	3.0	0-19	0.990	0.387
DSPA-R	34.82	22.39	35.0	0-79	-0.062	-.931
Word Identification	13.81	10.95	12.5	0-48	1.124	0.988
Word Attack	6.67	4.18	7.0	0-24	0.823	2.11

Note. ISF= Initial Sound Fluency; SDT=Sound Deletion Task; DSPA-R= Dynamic Screening of Phonological Awareness-Revised

Next, distributional characteristics of all variables from Sample 2 were presented in Tables 3.3 and 3.4. (Note that the SDT was not administered to students in Sample 2.) Inspection of the data revealed skewness statistics close to zero for the DSPA-R (.043 and -.145), similar to those noted in Sample 1. However, the distributional characteristics of the ISF differed from those seen in Sample 1. In both the full and the reduced Sample 2, the ISF was associated with a skewness statistic approaching or over a value of 1.0, indicating a floor effect. Figure 3.2 shows histograms for the predictor variables for the full Sample 2, and Figure 3.3. shows histograms for the reduced Sample 2.

Table 3.3*Descriptive statistics for the full Sample 2 (N=161)*

Measure	Mean	SD	Median	Range	Skew	Kurtosis
ISF	10.46	8.82	8.11	0-42	1.261	1.66
DSPA-R	36.63	25.75	34.0	0-78	0.043	-1.28
Word Identification	10.83	12.51	6.0	0-61	1.758	3.04
Word Attack	5.23	6.26	3.0	0-33	1.960	4.20

Note. ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness-Revised

Table 3.4*Descriptive statistics for the reduced Sample 2 (N=96)*

Measure	Mean	SD	Median	Range	Skew	Kurtosis
ISF	12.82	9.39	11.2	0-38	0.806	0.894
DSPA -R	44.59	26.26	48.0	0-78	-0.245	-1.05
Word Identification	14.12	14.41	8.0	0-61	1.303	1.17
Word Attack	6.74	7.10	6.0	0-33	1.679	3.35

Note. ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness-Revised

The distributions related to the Word Identification and Word Attack subtests were also positively skewed across all samples. Whereas many students were reading words and nonwords by the end of kindergarten, a sizeable number in each sample were still performing at the floor of the Word Identification and the Word Attack measures. Such a distribution of scores is not optimal for an outcome measure in a study examining the predictive validity of screening measures. However, such performance on the WRMT-R/NU is likely an accurate reflection of end of kindergarten reading achievement, and thus, are the outcome data this type of screening measure seeks to predict.

The Predictive Validity of DSPA-R as Compared to the SDT

The next set of analyses addressed the following research question: *Does a dynamic screening measure of phonological awareness administered in kindergarten add significantly to the prediction of reading achievement over and above a comparable static measure of phonological awareness utilizing the same items?* To answer this research question, data from Sample 1 were utilized. First, correlations among the variables were inspected. Second, hierarchical linear regression was conducted to examine the amount of unique variance accounted for by the DSPA-R over and above the SDT. Logistic regression was then conducted to determine the accuracy of the DSPA-R in predicting students' end of kindergarten reading outcomes as compared to the SDT. For all analyses, the DSPA-R and the SDT served as predictor variables. The outcome variables were Word Identification and Word Attack subtests from the WRMT-R/NU administered at the end of kindergarten. The outcome variables were treated both as continuous variables (correlations and linear regression) and dichotomized in terms of the presence or absence of RD. For all analyses, the focus was on investigating the extent to which the DSPA-R adds to the predictive validity of end of kindergarten reading outcomes over and above the SDT.

Correlations among the measures were first examined. As seen in Table 3.5, the correlation between the DSPA-R and the SDT was high. Because of the emphasis on prediction in this study, correlations between the predictors (i.e., the DSPA-R and the SDT) and the performance on outcomes (i.e., Word Identification and Word Attack) were of particular interest. The DSPA-R shared moderate correlations with

both Word Identification (.516) and Word Attack (.477). The SDT shows a moderate correlation with Word Identification (.486) but a reduced correlation with the Word Attack (.377).

Table 3.5
Correlations between predictors and outcome measures in Sample 1 (N=90)

Measure	1.	2.	3.
1. SDT			
2. DSPA-R	.840		
3. Word Identification	.486	.516	
4. Word Attack	.377	.477	.764

Note. SDT= Static Deletion Task; DSPA-R=Dynamic Screening of Phonological Awareness-Revised

Next, hierarchical linear regression was employed to investigate if the DSPA-R explained variance over and above the SDT in the prediction of end of kindergarten performance in Word Identification and Word Attack. To address this research question, the SDT was entered first into a regression model; then, the DSPA-R was entered.

As shown in Table 3.6, Model 1 included Word Identification as the outcome measure, and Model 2 included Word Attack as the outcome measure. The results of Model 1 indicated that the full model was significant, $R = .53$, $F(2, 87) = 16.55$, $p < .01$. Model 1 accounted for 28% variance in end of the year Word Identification performance. The DSPA-R accounted for a significant amount of variance (4%) over and above that of the SDT. The results of Model 2 indicated that the full model was significant, $R = .48$, $F(2, 87) = 13.81$, $p < .01$. Model 2 accounted for 23% of the variance in end of year Word Attack performance. The DSPA-R accounted for a

significant amount of variance (9%) over and above that of the SDT.

Table 3.6

Hierarchical linear regression analyses for predicting Word Identification and Word Attack in Sample 1 (N=90)

Variable	Word Identification			Word Attack		
	R^2	$R^2\Delta$	Semi-partial	R^2	$R^2\Delta$	Semi-partial
Model 1				Model 2		
SDT	.24		.10	.14		-.05
DSPA-R	.28	.04*	.20	.23	.09**	.30

Note. * $p < .05$; ** $p < .01$; SDT=Static Deletion Task; DSPA-R=Dynamic Screening of Phonological Awareness-Revised

Next, logistic regression was employed to examine the extent to which the SDT and the DSPA-R, alone or in combination, predicted students' reading outcomes. It is important to note that in educational settings, students are not typically identified with RD in kindergarten. Such a designation is not generally given until first or second grade. However, for the purposes of this research study, students who scored below a designated cutoff score at the end of kindergarten were labeled as RD. The Word Identification and Word Attack subtests of the WRMT-R/NU were chosen to serve as outcome measures because these or similar indices (e.g., Woodcock-Johnson Psychoeducational Battery-Revised; Woodcock & Johnson, 1989) are commonly used in educational settings to identify children with RD. In all logistic regression analyses, students were identified as RD if their score on a reading achievement outcome measure at the end of kindergarten was at or below the 25th percentile. The 25th percentile has been frequently used in reading research literature as a definition for RD, especially in the early elementary grades (e.g., Francis,

Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Siegel, 1992; Snow, Burns, & Griffin, 1998; Stanovich & Siegel, 1994). Furthermore, it was necessary to use sample statistics to identify the 25th percentile rather than the normative data provided in the WRMT-R/NU manual. This manual was revised over ten years ago, prior to the influence of the No Child Left Behind Act (NCLB; 2001). Although this legislation was formally confined to the upper elementary grades, instructional requirements associated with the new standards have impacted kindergarten curriculum. As a result, many kindergarten students are reading at levels well beyond those reported in the normative data provided by the WRMT-R/NU. Therefore, for this research study, cutoff scores for outcome measures were based on sample characteristics. These cutoff scores identified 28 and 27 students with RD based on end of kindergarten Word Identification and Word Attack, respectively.

In the logistic analyses, the SDT and the DSPA-R were first entered separately as single predictors, and then entered together in a sequential manner, the SDT followed by the DSPA-R. Classification information obtained from these logistic regressions was used to plot a ROC curve for each predictor as well as the combination of predictors. Recall that AUC refers to the area under a ROC curve, and values of AUC range from .5 (i.e., chance level) to 1.0 (i.e., perfect classification). The AUCs can be subjected to a rough guide of acceptability. Generally, values above .70 are considered to be “fair” and values above .80 are considered to be “good.”

Models 1-3 included data concerning end of kindergarten Word Identification.

In Model 1, SDT was entered as a single predictor. In Model 2, the DSPA-R was entered as a single predictor. In Model 3, SDT and DSPA-R were entered in a sequential fashion. As seen in Table 3.7, results of Model 1 showed the SDT was not a significant predictor by itself ($p=.061$). In Model 2, the DSPA-R was a significant predictor by itself ($p=.007$). In Model 3, with SDT entered first and the DSPA-R entered second, the SDT did not reach significance ($p=.503$) but the DSPA-R did reach significance ($p=.040$). Next, the results for predicting RD based on end of kindergarten Word Attack scores (Models 4-6) were examined. As in the above set of analyses, in Model 4, SDT was entered as a single predictor, and in Model 5, the DSPA-R was entered as a single predictor. In Model 6, SDT and DSPA-R were entered in a sequential fashion. In Model 4, the SDT was a significant predictor by itself ($p=.004$), and in Model 5, the DSPA-R was a significant predictor by itself ($p=.001$). In Model 6, with SDT entered first and the DSPA-R entered second, the SDT dropped as a significant predictor ($p=.907$) but the DSPA-R maintained as a significant predictor ($p=.014$). Note also that the DSPA-R had a stronger influence when predicting to Word Attack outcomes than to Word Identification outcomes.

Table 3.7

Logistic regression analyses predicting Word Identification and Word Attack in Sample 1 (N=90)

Measure	B	SE	Wald	<i>p</i> level	AUC
Word ID					
Model 1: SDT	-.092	.049	3.51	.061	.605
Model 2: DSPA-R	-.031	.011	7.37	.007	.667
Model 3: SDT	.059	.088	0.44	.503	
DSPA-R	-.042	.020	4.24	.040	.692
Word Attack					
Model 4: SDT	-.182	.063	8.37	.004	.635
Model 5: DSPA-R	-.050	.013	14.55	.001	.766
Model 6: SDT	.012	.099	0.014	.907	
DSPA-R	-.052	.021	5.99	.014	.772

Note. SDT= Static Deletion Task; DSPA-R= Dynamic Screening of Phonological Awareness-Revised; AUC= area under the curve

Using the above mentioned guidelines, the AUCs for predicting end of kindergarten Word Identification (Models 1-3) were less than adequate. The AUCs for Model 4, with SDT for predicting end of year Word Attack, was also less than adequate (.635); however, the AUCs associated with Model 5 and Model 6 revealed fair classification ability. The AUCs were then compared using a procedure developed by Hanley and McNeil (1983). This procedure accounts for the correlation introduced between the two AUCs as a result of using the same sample for each model. The value obtained from the formula provides evidence that the ROC curves are different if $z \geq 1.96$. When predicting to Word Identification, there were no significant differences between AUCs ($z = .37- .98$). When predicting to Word Attack, there were statistically significant differences in AUC between the SDT and DSPA-R ($z=2.19$) and the SDT and the combined screening ($z=2.25$). No difference was found between the DSPA-R and the combined screening ($z=.11$).

In sum, results of the above series of analyses showed that the DSPA-R outperformed the SDT as a predictor of reading achievement. Linear regression showed that the DSPA-R accounted for unique variance over and above the SDT. Additionally, logistic regression analyses indicated the DSPA-R, when entered after the SDT, significantly predicted reading outcomes. The SDT, however, was not a significant predictor in logistic regression models that also included the DSPA-R.

DSPA-R as Compared to the ISF

Next, the following research question was addressed: *Does a dynamic screening measure of phonological awareness administered in kindergarten add significantly to the prediction of reading achievement over and above a commonly used static measure of phonological awareness?* In this set of analyses, the predictive validity of the DSPA-R was examined. First, correlations among variables were inspected, and then hierarchical linear regression was employed. Finally, logistic regression was conducted. For all analyses, the focus was on investigating to what extent the DSPA-R adds to the predictive validity over and above the ISF. In analyses that included both variables, the ISF was entered first, followed by the DSPA-R. This order of entry was due to the interest in what the DSPA-R added to prediction of reading achievement over that predicted from a common static screening measure (i.e., the ISF). Data from all 251 students were utilized, with analyses completed separately for Sample 1, the full Sample 2, and the reduced Sample 2. The outcome variables were end of kindergarten Word Identification and Word Attack scores.

As displayed in Table 3.8, the correlations between the DSPA-R and the ISF were moderate, with higher correlations noted with the full and the reduced Sample 2 (.553 and .591) than in Sample 1 (.397). An inspection of the correlations between the DSPA-R and the Word Identification and Word Attack outcome measures showed moderate correlations across samples. Similar results were noted between the ISF and the outcome measures.

Table 3.8
Correlations between predictors and outcome measures

Sample 1 (N=90)	1.	2.	3.
1. ISF			
2. DSPA-R	.397		
3. Word Identification	.491	.516	
4. Word Attack	.511	.477	.764
<hr/>			
The full Sample 2 (N=161)			
1. ISF			
2. DSPA-R	.553		
3. Word Identification	.530	.444	
4. Word Attack	.495	.461	.869
<hr/>			
Reduced Sample 2 (N=96)			
1. ISF			
2. DSPA-R	.591		
3. Word Identification	.537	.426	
4. Word Attack	.531	.485	.875

Note. ISF=Initial Sound Fluency; DSPA-R=Dynamic Screening of Phonological Awareness-Revised

Next, a series of hierarchical linear regression analyses examined the amount of variance the DSPA-R accounted for over and above the ISF when predicting end of kindergarten reading outcomes as indexed by the Word Identification and Word Attack. In the following analyses, the variables were entered in a sequential manner, with the ISF entered first and the DSPA-R entered next.

First, results from Sample 1 were presented. As presented in Table 3.9, Model 1 included Word Identification as the outcome measure, and Model 2 included Word Attack as the outcome measure. The results of Model 1 indicated that the full model was significant, $R = .61$, $F(2, 87) = 25.40$, $p < .001$. Model 1 accounted for 37% variance in end of kindergarten Word Identification performance. The DSPA-R accounted for a significant amount of unique variance (13%) over and above the ISF. The results of Model 2 indicated that the full model was significant, $R = .60$, $F(2, 87) = 23.96$, $p < .001$. Model 2 accounted for 36% of the variance in end of kindergarten Word Attack performance, with the DSPA-R accounting for a significant amount of unique variance (9%) over and above the ISF.

Table 3.9
Hierarchical regression analyses for predicting Word Identification and Word Attack in Sample 1 (N=90)

Variable	Word Identification			Word Attack		
	R^2	$R^2\Delta$	Semi-partial	R^2	$R^2\Delta$	Semi-partial
Model 1				Model 2		
ISF	.24		.32	.26		.36
DSPA-R	.37	.13**	.36	.36	.09**	.31

Note. ** $p < .001$; ISF= Initial Sound Fluency measure of the DIBELS; DSPA-R= Dynamic Screening of Phonological Awareness-Revised

Next, data from the full Sample 2 were analyzed. As presented in Table 3.10, Model 1 included Word Identification as the outcome measure, and Model 2 included Word Attack as the outcome measure. The results of Model 1 indicated that the full model was significant, $R = .56$, $F(2, 157) = 36.15$, $p < .001$. This model, including both predictors, accounted for 31% variance in end of kindergarten Word Identification

performance. The DSPA-R accounted for a significant amount of unique variance (3%) over and above the ISF. The results of Model 2 indicated that the full model was significant, $R = .54$, $F(2, 157) = 33.16$, $p < .001$. Model 2 accounted for 32% of the variance in end of kindergarten Word Attack performance, with the DSPA-R accounting for a significant amount of unique variance (4%) over and above the ISF.

Table 3.10
Hierarchical regression analyses for predicting Word Identification and Word Attack in the full Sample 2 (N=161)

Variable	Word Identification			Word Attack		
	R^2	$R^2\Delta$	Semi-partial	R^2	$R^2\Delta$	Semi-partial
Model 1				Model 2		
ISF	.28		.34	.29		.34
DSPA-R	.31	.03*	.18	.32	.04*	.19

Note. * $p < .05$; ISF=Initial Sound Fluency; DSPA-R=Dynamic Screening of Phonological Awareness-Revised

Finally, data from the reduced Sample 2 were entered into linear regression analyses. As in the previous two analyses, Model 1 included Word Identification as the outcome measure, and Model 2 included Word Attack as the outcome measure. As presented in Table 3.11, the results of Model 1 indicated that the full model was significant, $R = .55$, $F(2, 93) = 20.54$, $p < .001$. Model 1 accounted for 31% variance in end of the year Word Identification performance. The DSPA-R accounted for a nonsignificant amount of variance (2%) over and above the ISF. The results of Model 2 indicated that the full model was significant, $R = .57$, $F(2, 93) = 22.63$, $p < .001$. Model 2 accounted for 33% of the variance in end of year Word Attack performance. The DSPA-R accounted for a significant amount of variance (5%) over

and above that of the ISF.

Table 3.11

Hierarchical regression analyses for predicting Word Identification and Word Attack in the reduced Sample 2 (N=96)

Variable	Word Identification			Word Attack		
	R^2	$R^2\Delta$	Semi-partial	R^2	$R^2\Delta$	Semi-partial
Model 1				Model 2		
ISF	.28		.35	.28		.30
DSPA-R	.31	.02	.16	.33	.05*	.21

Note. * $p < .05$; ** $p < .01$; ISF=Initial Sound Fluency; DSPA-R=Dynamic Screening of Phonological Awareness-Revised

Next, logistic regression was employed to examine the extent to which the predictors (ISF and DSPA-R), alone or in combination, predicted students' risk of being identified with RD at the end of kindergarten. As in the above analyses, the Word Identification and Word Attack subtests of the WRMT-R/NU were used as the two outcome measures. Students were identified as having RD if their score on an outcome measure was at or below the sample 25th percentile. For all analyses utilizing Sample 2, the cutoff score was associated with the 25th percentile of the reduced sample. For the full Sample 2, this identified 56 and 62 students with RD based on end of kindergarten Word Identification and Word Attack. For the reduced Sample 2, this identified 26 and 28 students with RD based on end of kindergarten Word Identification and Word Attack.

A series of logistic regression analyses were conducted to examine the relationship between the predictor variables (ISF and DSPA-R) and end of kindergarten reading outcomes (Word Identification and Word Attack). For each

sample, the first two logistic regressions included the ISF (Model 1) and the DSPA-R (Model 2) as individual predictors of Word Identification performance. The next logistic regression (Model 3) included ISF then DSPA-R, entered sequentially, to predict Word Identification performance. Models 4-6 included the same series of analyses predicting Word Attack performance.

The first set of analyses utilized data from Sample 1. As seen in Table 3.12, the DSPA-R in Model 3 did not improve the classification accuracy over and above the ISF ($p=.085$). Next, analyses predicted to end of the year Word Attack performance. Word Identification was regressed on the ISF alone (Model 4), the DSPA-R alone (Model 5), and on both predictors when entered in a sequential fashion (Model 6). The DSPA-R significantly improved the classification accuracy ($p=.003$) when entered after the ISF (Model 6). The AUCs associated with the DSPA-R alone were higher than the AUCs associated with the ISF alone. As was expected, the AUCs associated with the combined models were the highest for both Word Identification and Word Attack. However, these differences in AUCs across models were not significant ($z=.26-1.66$).

Table 3.12*Classification indices for the logistic regression analyses for Sample 1 (N=90)*

Outcome measure	B	SE	Wald	<i>p</i>	AUC
Word ID					
Model 1: ISF	-.105	.032	10.414	.001	.721
Model 2: DSPA-R	-.031	.011	7.373	.007	.667
Model 3: ISF	-.089	.034	6.794	.009	
DSPA-R	-.021	.012	2.972	.085	.749
Word Attack					
Model 4: ISF	-.123	.035	12.426	.001	.751
Model 5: DSPA-R	-.050	.013	14.553	.001	.766
Model 6: ISF	-.101	.038	7.093	.008	
DSPA-R	-.043	.014	9.064	.003	.821

Note. ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness

The next series of analyses utilized data from the full Sample 2. First, logistic regression analyses were conducted with Word Identification as the outcome measure. As shown in Table 3.13, Word Identification was regressed on the ISF alone (Model 1), the DSPA-R alone (Model 2), and on both predictors when entered in a sequential fashion (Model 3). As noted in Model 3, the DSPA-R improved the classification accuracy significantly ($p=.001$) when added after the ISF. Next, analyses predicted to end of the year Word Attack performance. Word Attack was regressed on the ISF alone (Model 4), the DSPA-R alone (Model 5), and on both predictors when entered in a sequential fashion (Model 6). The DSPA-R significantly improved the classification accuracy ($p=.001$) over and above the ISF. The AUCs associated with the DSPA-R alone were higher than the AUCs associated with the ISF alone. As was expected, the AUCs associated with the combined models were the highest for both Word Identification and Word Attack. When predicting to Word Identification, there was a significant differences in AUCs between the ISF and the

combined measure ($z=1.987$). There were not significant differences in AUC model indices between the ISF and the DSPA-R ($z=1.59$) or the DSPA-R and the combined measures ($z=.32$). When predicting to Word Attack, there were no significant differences between AUCs ($z= .71-1.40$).

Table 3.13
Classification Indices for the logistic regression analyses for the full Sample 2 (N=161)

Outcome measure	B	SE	Wald	<i>p</i>	AUC
Word Identification					
Model 1: ISF	-.113	.029	15.682	.001	.709
Model 2: DSPA-R	-.044	.008	28.353	.001	.774
Model 3: ISF	-.064	.032	3.915	.048	
DSPA-R	-.037	.009	16.803	.001	.786
Word Attack					
Model 4: ISF	-.137	.030	20.593	.001	.735
Model 5: DSPA-R	-.041	.008	27.042	.001	.762
Model 6: ISF	-.098	.033	8.697	.003	
DSPA-R	-.030	.009	12.170	.001	.789

Note. Word ID= ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness; AUC= area under the curve

Finally, data from the reduced Sample 2 were analyzed and presented in Table 3.14. Logistic regression analyses were conducted with Word Identification as the outcome measure. Word Identification was regressed on the ISF alone (Model 1), the DSPA-R alone (Model 2), and on both predictors when entered in a sequential fashion (Model 3). As noted in Model 3, the DSPA-R improved the classification accuracy significantly ($p=.045$) when added after the ISF; note also that when both predictors were included in this model, the ISF dropped out of the model as a significant predictor ($p=.064$). Next, analyses predicted to end of the year Word

Attack performance. Word Attack was regressed on the ISF alone (Model 4), the DSPA-R alone (Model 5), and on both predictors when entered in a sequential fashion (Model 6). The DSPA-R significantly improved the classification accuracy ($p=.016$) over and above the ISF. The AUCs associated with the DSPA-R alone were higher than the AUCs associated with the ISF alone. As was expected, the AUCs associated with the combined models were the highest for both Word Identification and Word Attack. However, none of these differences between AUCs were statistically significant ($z=.533-.705$).

Table 3.14
Classification Indices for the logistic regression analyses for the reduced Sample 2 (N=96)

Outcome measure	B	SE	Wald	<i>p</i>	AUC
Word ID					
Model 1: ISF	-.114	.036	9.980	.002	.742
Model 2: DSPA-R	-.034	.010	11.878	.001	.755
Model 3: ISF	-.075	.040	3.425	.064	
DSPA-R	-.022	.012	3.692	.045	.760
Word Attack					
Model 4: ISF	-.167	.043	15.109	.001	.796
Model 5: DSPA-R	-.044	.011	17.704	.001	.801
Model 6: ISF	-.117	.046	6.499	.011	
DSPA-R	-.029	.012	5.824	.016	.825

Note. ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness; AUC= area under the curve

Results from the above series of analyses supported the previous findings that in most cases, the DSPA-R added significantly to the prediction of reading outcomes. All but one of the AUCs associated with the DSPA-R were associated with fair or good classification accuracy; the exception was in Sample 1 predicting to end of year Word Identification. Additionally, the DSPA-R was a stronger predictor of Word

Attack reading outcomes than of Word Identification outcomes.

The DSPA-R as a Supplemental Screening Measure

The DSPA-R was developed in part to serve as a supplemental measure in a two-step identification process in order to reduce false positives associated with many screening measures. To assess its usefulness in this regard, additional analyses were carried out. In these analyses, an initial detection of risk status based on the ISF was established. The DSPA-R was then used as a supplemental screening measure.

Classifications were categorized in the following manner: (1) True Positives- poor readers identified as at risk; (2) True Negatives- good readers identified as not at risk; (3) False Positives- good readers identified as at risk; and (4) False Negatives- poor readers identified as not at risk. Sensitivity and specificity of prediction were then used as indices in the evaluation of this screening approach.

As used in universal screenings, the developers of the ISF recommend an early kindergarten cutoff score of 8 to identify students at risk. In the first set of analyses, the ISF variable was dichotomized, with a score of less than 8 associated with at risk and a score of 8 or above associated with not at risk. The DSPA-R was then used as a supplemental screening measure to re-classify students deemed to be at risk by the ISF. The mean DSPA-R score from the sample was used as the cutoff score. For Sample 1, this value was 35. Again, the mean value from the reduced Sample 2 (45) was used for all analyses utilizing data from Sample 2. Other cutoff scores on the DSPA-R were also investigated but the mean score provided the best trade-off between sensitivity and specificity.

In the next series of analyses, the classification accuracy of the dichotomized ISF variable was examined, with the outcome variable as end of the kindergarten Word Identification performance (Model 1). The next step was to investigate how the classification rates were affected by using the DSPA-R. Model 2 included both the ISF and the DSPA-R, administered in a sequential fashion. The same analyses were replicated when predicting to end of year Word Attack performance (Models 3 and 4).

Data from Sample 1 are presented first. As seen in Table 3.15, results indicated that using the ISF recommended cutoff score as the sole predictor resulted in good specificity but very poor sensitivity. Out of the 28 students with RD based on Word Identification scores at the end of kindergarten, only 9 were identified as at risk by the ISF. Similarly, out of the 27 students with RD based on Word Attack scores at the end of kindergarten, 9 were identified as at risk by the ISF. This model also resulted in very few false positives (4) left for the DSPA-R to correctly identify when entered into the regression (Models 2 and 4). Thus, a supplemental model did not have the opportunity to add significantly to the prediction of reading outcomes when the recommended ISF cutoff score was used.

Table 3.15

Classification indices across models utilizing the ISF alone and the ISF with the DSPA-R as a supplemental screening measure in Sample 1 (N=90)

	TP	TN	FP	FN	Specificity	Sensitivity
Word Identification						
Model 1: ISF < 8	9	58	4	19	.94	.33
Model 2: ISF + DSPA-R ≤ 35	8	59	3	20	.95	.29
Word Attack						
Model 3: ISF < 8	9	59	4	18	.95	.33
Model 4: ISF + DSPA-R ≤ 35	8	60	3	19	.95	.30

Note. ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness

This same two-step approach was conducted with data from the full and the reduced Sample 2, as seen in Tables 3.16 and 3.17. Using a cutoff score of 8 for the ISF resulted in poor specificity and sensitivity in both the full and the reduced Sample 2. Whereas the DSPA-R did reduce false positives, the overall results were limited by the sensitivity of the ISF.

Table 3.16

Classification indices across models utilizing the ISF alone and the ISF with the DSPA-R as a supplemental screening measure in the full Sample 2 (N=161)

	TP	TN	FP	FN	Specificity	Sensitivity
Word Identification						
Model 1: ISF < 8	38	67	38	18	.64	.68
Model 2: ISF + DSPA-R ≤ 45	36	80	25	20	.76	.65
Word Attack						
Model 3: ISF < 8	43	66	33	19	.67	.69
Model 4: ISF + DSPA-R ≤ 45	39	77	22	23	.78	.63

Note. ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness

Table 3.17

Classification indices across models utilizing ISF alone and the ISF with the DSPA-R as a supplemental screening measure in the reduced Sample 2 (N=96)

	TP	TN	FP	FN	Specificity	Sensitivity
Word Identification						
Model 1: ISF < 8	16	53	17	10	.77	.62
Model 2: ISF + DSPA-R ≤ 45	15	59	11	11	.84	.58
Word Attack						
Model 3: ISF < 8	17	52	16	11	.76	.61
Model 4: ISF + DSPA-R ≤ 45	16	58	10	12	.85	.57

Note. ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness

Additional Uses of the DSPA-R as a Supplemental Screening Measure

In the above analyses, the overall sensitivity was influenced considerably by the use of the recommended cutoff score for the ISF. False negatives were not included in the secondary analysis, and thus, there was no way to increase the sensitivity of the DSPA-R as a supplemental screening measure. This was particularly true for Sample 1.

Recall that Jenkins (2003) suggested that the best way to judge the classification accuracy of a screening measure is to choose a high sensitivity level and then evaluate the specificity/false positive rate. Using a two-step process, the first screening measure can set a cutoff score associated with a high sensitivity level. Then, the cutoff score for the supplemental screening can be set to reduce the number of false positives while maintaining an adequate level of sensitivity.

This procedure was followed for the next set of analyses. First, a series of classification indices were investigated for each sample to identify a cutoff score on the ISF that would result in sensitivity levels above .90. Next, corresponding DSPA-R cutoff scores were identified that maintained a high level of sensitivity while showing the highest reduction in the number of false positives.

These analyses were first completed with data from Sample 1; see Table 3.18. Results of the analyses showed that when predicting to Word Identification, an ISF cutoff score of 26 yielded a sensitivity of .92 and a specificity of .18 (Model 1). Although this cutoff score resulted in high sensitivity levels, there were many false positives (51). However, information from the DSPA-R decreased this number of

false positives to 36, while maintaining an adequate level of sensitivity (.82; Model 2). Similar results were noted for predicting end of kindergarten Word Attack (see Models 3 and 4).

Table 3.18

Classification indices in two-step process maintaining sensitivity above .90 and reducing false positives in Sample 1 (N=90)

	TP	TN	FP	FN	Specificity	Sensitivity
Word Identification						
Model 1 (ISF <26)	26	11	51	2	.18	.92
Model 2 (DSPAR < 46)	23	26	36	5	.42	.82
Word Attack						
Model 3 (ISF<26)	25	14	49	2	.22	.92
Model 4 (DSPAR<45)	24	32	31	3	.51	.88

Note. ISF= Initial Sound Fluency; DSPAR= Dynamic Screening of Phonological Awareness

The same approach was used with the full and the reduced Sample 2, as shown in Tables 3.19 and 3.20. In the full Sample 2, when identifying end of kindergarten Word Identification, an ISF cutoff score of 14 yielded a sensitivity of .92 and a specificity of .36. As discussed above, the use of these cutoff scores resulted in a large number of students misidentified with RD (67). However, information from the DSPAR decreased false positives to 43, while still maintaining an adequate level of sensitivity. When identifying RD for end of kindergarten based on Word Attack in Sample 2, an ISF cutoff score of 13 yielded a sensitivity of .93 and a specificity of

.44. Fifty-five students were misidentified with RD, but information from the DSPA-R decreased false positives to 39. Similar results were noted in the reduced Sample 2.

Table 3.19

Classification indices in two-step process maintaining sensitivity above .90 and reducing false positives in the full Sample 2 (N=161)

	TP	TN	FP	FN	Specificity	Sensitivity
Word Identification						
Model 1 (ISF <14)	52	38	67	4	.36	.92
Model 2 (DSPA-R < 45)	48	62	43	8	.59	.86
Word Attack						
Model 3 (ISF <13)	58	44	55	4	.44	.93
Model 4 (DSPA ≤ 48)	52	60	39	10	.61	.84

Note. ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness

Table 3.20

Classification indices in two-step process maintaining sensitivity above .90 and reducing false positives in the reduced Sample 2 (N=96)

	TP	TN	FP	FN	Specificity	Sensitivity
Word Identification						
Model 1 (ISF <20)	24	19	51	2	.27	.92
Model 2 (DSPA-R ≤ 50)	22	42	28	4	.60	.85
Word Attack						
Model 3 (ISF <13)	26	39	29	2	.57	.92
Model 4 (DSPA <55)	24	46	22	4	.67	.86

Note. ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness

The above analyses provided initial evidence for a practical application of using the ISF and the DSPA-R in a two-step screening process. However, it is important to investigate whether this approach would result in higher classification indices than those associated with administering all students both measures, or each individually. To investigate this possibility, the specificity values associated with the supplemental models (Tables 3.18-3.20) were compared to those of the combined screening, the DSPA-R alone, and the ISF alone (Tables 3.7-3.9), when sensitivity was held constant to that found with the supplemental model.

As noted in Table 3.21, results from Sample 1 showed that the identification methods were fairly equivocal. One exception was noted when predicting end of kindergarten Word Attack performance using the ISF; in this case, the specificity dropped to 30. Different results were noted in Sample 2. For the full Sample 2, the supplemental approach yielded the highest specificity level. For the reduced Sample 2, the supplemental approach showed the highest specificity when predicting Word Attack outcome. However, when predicting end of kindergarten Word Identification performance, the DSPA-R alone showed the highest specificity level (.63). These results provide preliminary support that the supplemental approach, or perhaps the DSPA-R alone, may be the best choice for a screening measure.

Table 3.21

Specificity for the supplemental model, the combined model, the DSPA alone, and the ISF alone when sensitivity is held constant to that found with the supplemental model

	Sensitivity		Specificity		
		Supplemental	Combined	DSPA-R	ISF
Sample 1:					
Word ID	.82	.42	.50	.50	.50
Word Attack	.88	.51	.51	.51	.30
Sample 2 (full):					
Word ID	.86	.59	.56	.55	.53
Word Attack	.84	.61	.50	.52	.53
Sample 2 (red.):					
Word ID	.85	.60	.49	.63	.30
Word Attack	.86	.67	.60	.63	.59

Note. ISF= Initial Sound Fluency; DSPA-R= Dynamic Screening of Phonological Awareness

CHAPTER IV: GENERAL DISCUSSION

The purpose of this research was to investigate the validity and usefulness of a dynamic screening of phonological awareness with students at the beginning of kindergarten. Four research questions were addressed. The first question concerned the reliability of a dynamic screening measure. In an RTI framework, universal screening is utilized to inform placement of students into supplemental reading instruction, and thus, it is crucial that measures used in this process have adequate reliability. The extant literature on dynamic assessment has not typically reported the reliability of the measures (i.e., Caffrey, 2006; Ferrara et al., 1986). This is likely due to the nature of treatment-oriented dynamic assessments, which are typically lengthy and lack standardization (e.g., Budoff, 1974; Feuerstein, 1979; Feuerstein & Rand, 1974). However, the standardized nature of the graduated prompts methodology in this study allowed for the measurement of reliability estimates. In Study I, the DSPA was administered by school personnel with reported limited knowledge of dynamic assessment. Despite their limited knowledge, after a two-hour training session they demonstrated high fidelity of administration (.90). Furthermore, the inter-rater (.98) and test-retest (.86) reliabilities were quite high and easily meet the .80 acceptability level proposed by Salvia and Ysseldyke (1988). Therefore, Study 1 contributes to the literature by showing that a dynamic screening measure of phonological awareness can be administered with high reliability and fidelity.

The second research question was concerned with the distributional characteristics of the dynamic screening measure. Past research has shown that young

children often perform at the floor on phonological awareness measures (e.g., Catts et al., 2009; Spector, 1992; Rathvon, 2004). Many children who perform poorly on such measures may not be truly at risk for RD; instead, their poor performance may be a reflection of their limited exposure to literacy or lack of understanding of task requirements. The initial version of the dynamic screening measure (i.e., the DSPA) was designed to reduce the floor effects seen in static phonological awareness measures by providing feedback/instruction when needed. Despite the feedback/instruction, the findings from Study I showed this version had a positively skewed distribution, which suggested a floor effect present in the data. However, skewness statistics showed that the SDT and the ISF had even more of a floor effect than the DSPA. Thus, a revised version of the dynamic screening measure (i.e., the DSPA-R) was developed. In Study II this revised measure was administered to three samples of kindergarten students. The distributional characteristics of the DSPA-R were compared to those of the SDT and the ISF. An inspection of the data associated with the DSPA-R revealed skewness statistics close to zero, indicating the absence of floor effects. This differed from the skewness statistic associated with the SDT, which indicated a floor effect associated with the data. The distribution of scores from the ISF in Sample 1 showed no positive skew. However, the distributions of the ISF for the full and the reduced Sample 2 were similar to those seen in the SDT. These findings showed that the DSPA-R, when administered at the beginning of kindergarten, did not display the floor effects often observed in similar static screening measures.

The third research question concerned the predictive validity of the DSPA-R as compared to the SDT. Recall that the DSPA-R and the SDT were both syllable/phoneme deletion screening measures comprised of identical items. The sole difference between the two was the examiner feedback provided in the former. This comparison of the two measures allowed for the direct evaluation of the dynamic component of the screening measure. Results of hierarchical linear regression analyses showed that the DSPA-R accounted for a significant amount of variance over and above the SDT when predicting reading achievement. Additionally, logistic regression analyses showed that the DSPA-R significantly predicted reading outcomes. Furthermore, when the SDT and DSPA-R were sequentially entered into the same logistic regression model, the SDT dropped out as a significant predictor. These findings provided preliminary support that the dynamic nature of the DSPA-R improved the predictive ability of a static syllable/phoneme deletion task.

The final research question concerned a comparison between the DSPA-R and the ISF. The ISF was chosen for this comparison because of its widespread use as a universal screening measure in elementary schools. It is common practice for school districts to use ISF scores to inform placement of kindergarten students in Tier 2 intervention. Despite its use, recent research has shown that the ISF is associated with high false positive rates and limited prediction of RD (Catts et al., 2009). The DSPA-R, in part, was developed to reduce the number of false positives when used as a supplement to the ISF or other similar screening measures. Across three samples of

kindergarten students, results from hierarchical regression analyses showed that the DSPA-R accounted for a significant amount of variance over and above the ISF.

To further compare the DSPA-R and the ISF, logistic regression analyses were employed. These analyses examined the predictive validity of the DSPA-R by itself, combined with the ISF, and as a supplemental screening measure. Results indicated that, across the three samples, both the ISF and the DSPA-R were significant single predictors of reading outcomes. The DSPA-R also added significantly to the prediction of reading outcomes when entered in logistic regression models after the ISF. The one exception was noted in Sample 1, in which the DSPA-R did not improve upon the ability to predict Word Identification. In most cases, the AUCs associated with the DSPA-R as well as the ISF had fair classification accuracy. However, across samples, the classification accuracy for the combination of the ISF and DSPA-R were fair to good, with AUCs ranging from .749-.825. Also note that the predictive validity was generally better when predicting to Word Attack outcomes than to Word Identification across all analyses. This is expected given that phonological awareness should be more directly related to phonological decoding than to sight word reading, the latter of which is the skill measured by the Word Identification subtest (Ehri, 1998).

The usefulness of the DSPA-R as a supplemental screening measure to the ISF was further examined by imposing cutoff scores in order to increase classification accuracy. In the first set of analyses, the cutoff score recommended by the DIBELS developers (Good & Kaminski, 2003) was used to identify poor readers, and then, the

DSPA-R was used to reclassify those students deemed to be at risk. In Sample 1, these analyses resulted in adequate specificity but very low sensitivity. In the full and the reduced Sample 2, the specificity values approached recommended levels but the sensitivity values continued to be poor. Additional follow-up analyses utilized a screening approach suggested by Jenkins (2003). These analyses used cutoff scores for the ISF that maximized sensitivity levels ($>.90$), and then cutoff scores for the DSPA-R were chosen to reduce the number of false positives while maintaining an acceptable number of false negatives. This procedure yielded acceptable sensitivity (.82-.88) levels when predicting to end of kindergarten across the samples, with accompanying specificity levels of .42-.67. Although these values are lower than the acceptable levels, they are quite consistent with an RTI approach. As noted previously, when choosing screening measures, greater emphasis should be placed on sensitivity to ensure identification of students at risk for RD. In the final analyses, the classification accuracies of the supplemental models were compared to those of models including combined screening, the DSPA-R alone, and the ISF alone. In Sample 1, the supplemental screening approach did not yield higher accuracy rates than the combined screening or the DSPA-R or the ISF alone. In Sample 2, while the DSPA-R alone maintained good levels of specificity, the supplemental approach did better in 3 of the 4 models. Although the results from this research are preliminary, they suggest that the DSPA-R alone or in combination with other screening measures is a valid and useful screening approach. It would be informative to conduct a cost-benefit analysis, in which the benefits associated with using the DSPA-R as a single

screening measure could be compared to its use as a supplemental measure to the ISF or a similar measure. Recently, Gresham (2002) advocated the use of such a cost-benefit analysis in determining financial costs to school districts in using RTI approaches.

A Dynamic Approach to Universal Screening

Past research has shown that performances on phonological awareness tasks are associated with later reading achievement, but young students often perform at the floor on such measures, thereby limiting the predictive abilities of such tasks when given to young children (e.g., Catts et al., 2009; Spector, 1992; Rathvon, 2004). The results of this research study provide preliminary evidence that adding examiner prompts/feedback to a phonological awareness task reduced floor effects, and in turn, increased the predictive validity of the measure when administered to kindergarten students. These findings converged with those of both Spector (1992) and O'Connor and Jenkins (1999) in that a dynamic measure improved prediction accuracy in young students over a similar static phonological awareness measure. It is likely that the feedback provided in the dynamic measures allowed students with partially-developed phonological awareness skills to separate themselves from those with little to no knowledge. This is particularly important at the beginning of kindergarten, because as noted previously, poor performance could be related to limited literacy experience. This hypothesis is supported by strong evidence that literacy experience and instruction leads to increased performance on phonological awareness measures

(e.g., Castles & Coltheart, 2004; Hogan, Catts, & Little, 2005; Mann & Wimmer, 2002).

Although this study was not designed to determine why a dynamic screening measure may be a more accurate predictor of later reading achievement than a similar static measure, it is possible that the work by Vygotsky (1978) can inform this discussion. Recall Vygotsky's notion of the zone of proximal distance, defined as the distance between the level of functioning students demonstrate independently and the higher level at which they function with adult scaffolding. The dynamic screening measure utilized in this research can be considered within this framework. The students who benefited the most from the prompts/feedback might be the students who also benefited the most from classroom instruction throughout kindergarten. Therefore, the dynamic assessment wasn't just a measure of individual variations in a phonological awareness task; it was a measure of individual variations in a student's ability to respond to adult instruction. In a comprehensive review of dynamic assessment, Grigorenko and Sternberg (1998) proposed that dynamic assessment taps student learning potential in a way that is distinct from static measures. Specifically, static measures typically assess already-developed abilities whereas dynamic measures are an indicator of a student's potential to learn new information.

This research also adds to the existing dynamic assessment literature by providing initial support for its use as a screening measure for risk for RD in kindergarten students. The dynamic assessments utilized by both Spector (1992) and O'Connor and Jenkins (1999) were predictive of later reading ability; however, they

were too lengthy to be administered in a universal screening approach. The DSPA-R was developed with universal screening in mind. Therefore, this study was one of the first to explore the use of a dynamic measure as part of a universal screening approach.

The findings converged with other research that has shown a multivariate screening approach may increase the accuracy of early identification (e.g., Catts et al., 2001; Compton et al., 2006; Foorman et al., 1998; O'Connor and Jenkins, 1999). Scarborough (1998) discussed the increase in predictive accuracy when researchers have combined kindergarten measures, rather than using a single variable, to predict later reading achievement. However, combining predictors also has its disadvantages. Although a large number of independent measures will most likely result in high prediction power, this comes with practical limitations related to cost and money. Therefore, a screening battery should predict later reading achievement accurately but not at the expense of the efficiency necessary for universal screening. As noted above, this research study found that a combination of two early kindergarten screening measures, the ISF and the DSPA-R, yielded acceptable sensitivity (.82-.88) levels when predicting reading outcomes. It is true that the specificity levels associated with this combination model were lower than desired (.42-.67). Although there are costs associated with false positives in an educational setting, these costs are much less worrisome than those associated with false negatives. Students incorrectly identified as not at risk for RD are not provided with supplemental instruction at a

young age, and as a result, do not experience the reading gains associated with early supplemental intervention.

The use of a dynamic screening measure as part of universal screening might be particularly beneficial with students at the beginning of formal schooling. As previously mentioned, many early predictors are associated with floor effects and an associated high rate of false positives. Researchers have suggested that universal screening measures might be more appropriate at the beginning of first grade than kindergarten because more accurate determination of risk for RD occurs as students experience more formal reading instruction (e.g., Fletcher et al., 2002; O'Connor & Jenkins, 1999; Torgesen, Burgess, Wagner, & Rashotte, 1996). However, recall that studies have shown that intervention provided in kindergarten resulted in at-risk students achieving normal reading proficiency in first grade and beyond (Cavanaugh et al., 2004; O'Connor et al., 2005; Scanlon et al., 2005; Simmons et al., 2008; Vellutino et al., 2006). Postponing universal screening results in students not being identified as at risk in kindergarten, and therefore, not receiving supplemental instruction until first grade. Furthermore, while O'Connor & Jenkins (1999) showed a substantial increase in predictive power when waiting until first grade, other researchers have shown that waiting until first grade resulted in minimal benefits related to classification accuracy (Torgesen et al., 1996; Foorman et al., 1998).

It is reasonable to suggest that a dynamic screening measure of phonological awareness might be particularly useful within an RTI framework. Most RTI models require students to remain in Tier 2 intervention for as many as 10-30 months before

being considered a “nonresponder.” This means that a large part of the school year could pass without those students receiving individualized intervention (Haager et al., 2007). A dynamic screening measure might serve to more quickly and/or accurately identify those students who will ultimately show poor response to Tier 2 intervention. If this is the case, students who perform poorly on a dynamic screening measure could receive individualized instruction more quickly, and thus, eliminate participating in many weeks of a Tier 2 intervention that might not be effective.

Limitations of the Current Research Study

It is important to acknowledge the limitations associated with this research. First, although logistic regression procedures are fairly robust to non-normal variable distributions (Tabachnick & Fidell, 1996), it is not known how the skewed distributions of the Word Identification and Word Attack measures affected the results of the predictive analyses. Outcome measures obtained from the end of first grade or beyond are more desirable for a prediction study, and therefore, these results should be interpreted cautiously until the results are replicated with such outcome data. Furthermore, only standardized tests were utilized as outcome measures. Good performance on such tests does not necessarily generalize to good performance in a classroom setting. It is possible that a curriculum-based outcome measure would provide additional insight into the predictive ability of a dynamic screening measure as it relates to classroom achievement.

Another limitation is related to the limited number of predictor variables utilized in the analyses. Although results from this research showed that the DSPA-R

improved identification accuracy over and above other phonological awareness measures, the measures used were limited in scope. Variables that tap other cognitive-linguistic aspects of reading achievement were not included in this research. Recall that, at best, a combination of the phonological awareness measures used in this research accounted for only 37% of the variance in reading achievement. It is possible that the predictive accuracy of the dynamic assessment might increase if additional screening measures are included.

Another limitation is the omission of information related to classroom instruction and/or additional intervention for students deemed to be at risk for RD. For many students in this study, there was some additional intervention between universal screening assessments and outcome measurements. The provision of intervention amongst the different schools (and classrooms) makes the interpretation of data concerning classification accuracy more difficult. As Good, Cummings, and Powell-Smith (2008) point out, intervention often improves the outcomes of at-risk children, and as a result, estimates classification accuracy are compromised. However, this problem is unavoidable when conducting research in an educational setting. Future investigations could address this problem by obtaining a large enough sample to utilize multilevel modeling techniques, which could take classroom and school effects into account.

Considerations for Future Research

Although the results of this research are promising for the use of the DSPA-R as an early screening measure, future research is warranted to both replicate and

extend the results of this investigation. Specifically, predictive validity should be examined in a larger sample obtained from a more diverse population. Additionally, variables included in the analyses should be expanded to include a wider range of predictor variables, such as more general language or cognitive measures, as well as a wider range of outcome measures, including both timed and untimed measures of word identification and decoding.

Additionally, future research should explore the instructional implications of the use of a dynamic assessment. As discussed by Sternberg and Grigorenko (2002), “in order for an educator to evaluate a student’s ability to learn, the educator needs to teach students something and then observe their learning.” This is essentially what occurs in a dynamic assessment. The prompting hierarchy provided in the DSPA-R may provide educators with ideas for instructional support necessary for a child to succeed. For example, one student might respond well to intonation cues, whereas another student might need the addition of a visual/motor cue (i.e., tapping out the syllables or phonemes) in order to provide the correct answer. This type of information is important for planning supplemental instruction found in Tier 2, and in fact, it is this type of information that more commonly utilized static screening measures are not able to provide. Further research could identify the prompts that are most salient to particular students and then further explore the utility of such prompts in a phonological training intervention study.

References

- Adams, M.J. (1990). *Beginning to read: Thinking and learning about print*. Urbana-Champaign, IL: Reading Research and Education Center.
- Badian, N.A. (1993). Predicting reading progress in children receiving special help. *Annals of Dyslexia*, 43, 90-109.
- Badian, N.A. (1994). Preschool prediction: Orthography and phonological skills, and reading. *Annals of Dyslexia*, 44, 3-25.
- Ball, E. W. (1993). Assessing phoneme awareness. *Language, Speech, & Hearing Services in the Schools*, 24, 130-139.
- Berkeley, S., Bender, W.N., Peaster, L.G., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities*, 42, 85-95.
- Blachman, B. (1989). Phonological awareness and word recognition: Assessment and intervention. In A. Kamhi and H. Catts (Eds.), *Reading disabilities: A developmental language perspective* (pp. 133-158). Boston, MA: College-Hill Press.
- Bowers, P. (1995). Tracing symbol naming speed's unique contributions to reading disabilities over time. *Reading and Writing: An Interdisciplinary Journal*, 7, 1-28.
- Bridges, M.S., & Catts, H.W. (2008). Dynamic Screening of Phonological Awareness. Pre-publication version. Moline, IL: Linguisystems.

- Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26*(3),429-438.
- Budoff, M. (1974). *Learning potential and educability among the educable mentally retarded*. (Final Report Project No. 312312). Cambridge, MA: Research Institute for Educational Problems, Cambridge Mental Health Association.
- Budoff, M. (1987). The validity of learning potential. In C.S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 52-81).
- Caffrey, E. (2006). *A comparison of dynamic assessment and progress monitoring in the prediction of reading achievement for students in kindergarten and first grade*. Unpublished doctoral dissertation, Vanderbilt University, Nashville.
- Carron, D.T., & Scott, K.G. (1992). Risk assessment in preschool children: Research implications for early detection of educational handicaps. *Topics in Early Childhood Special Education, 12*, 196-211.
- Camptone, J.C. (1989). Assisted assessment: A taxonomy of approaches and an outline of strengths and weakness. *Journal of Learning Disabilities, 22*, 151-165.
- Campione, J., & Brown, A. (1987). Linking dynamic assessment with school achievement. In C.S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82-115). New York: Guilford.

- Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*, 77-111.
- Catts, H.W. (1999). *Dynamic Assessment of Phonological Awareness*. Unpublished test.
- Catts, H.W., Fey, M.E., Zhang, X., & Tomblin, J.B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*, *32*, 38-50.
- Catts, H.W. & Kamhi, A.G., (2005). *Language and Reading Disabilities*. Boston: Allyn & Bacon.
- Catts, Hugh W., Petscher, Y., Schatschneider, C., Bridges, M.S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, *42*, 163-176.
- Cavanaugh, C. L., Kim, A., Wanzek, J., & Vaughn, S. (2004). Kindergarten reading interventions for at-risk students: Twenty years of research. *Learning Disabilities: A Contemporary Journal*, *2*, 9-21.
- Chafouleas, S.M., VanAuken, T., & Dunham, K. (2001). Not all phonemes are created equal: The effects of linguistic manipulations on phonological awareness tasks. *Journal of Psychoeducational Assessment*, *19*, 216-226.
- Compton, D.L., Fuchs, D., Fuchs, L.S., & Bryant, J.D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of

decision rules and procedures. *Journal of Educational Psychology*, 98, 394-409.

- Denton, C.A., & Mathes, P.G. (2003). Intervention for struggling readers: Possibilities and challenges. In B.R. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 229-251). Timonium, MD: York.
- Dickman, G.E. (2006). RTI and Reading: Response to Intervention in a Nutshell. *Perspectives on Language and Literacy, Special Conference Edition*. International Dyslexia Association: Baltimore, MD.
- Dollaghan, C.A. (2007). *The Handbook for Evidence-based Practice in Communication Disorders*. Baltimore, MD: Brookes.
- Duncan, S.E., & DeAvila, E.A. (1998). *Pre-LAS*. Monterey, CA: CTB/McGraw-Hill.
- Elbro, C., Borstrom, I., & Petersen, D.K. (1998). Predicting dyslexia from kindergarten: the importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly*, 33, 36-60.
- Ehri, L.C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J.L. Metsala & L.C. Ehri (Eds.), *Word recognition in literacy* (pp. 3-40). Mahwah, NJ: Lawrence Erlbaum Associates.
- Elkonin, D.B. (1973). USSR. In J.A. Downing (Ed.), *Comparative reading: Cross-national studies of behavior and processes in reading and writing* (2nd ed., pp. 551-579). New York: Macmillan.

- Felton, R.H., & Brown, I.S. (1990). Phonological predictors of specific reading skill in children at risk for reading failure. *Reading and Writing, 2*, 39-59.
- Ferrara, R.A., Brown, A.L., & Campione, J.C. (1986). Children's learning and transfer of inductive reasoning rules: Studies in proximal development. *Child Development, 57*, 1087-1099.
- Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. Baltimore, MD: University Park Press.
- Feuerstein, R., Haywood, H. C., Rand, Y., Hoffman, M. B., & Jensen, M. (1982). *Examiner manuals for the Learning Potential Assessment Device*. Jerusalem: Hadassah-WIZO-Canada Research Institute.
- Feuerstein, R., & Rand, J. (1974). Mediated learning experiences: An outline of proximal etiology for differential development of cognitive functions. *International Understanding, 9-10*, 7-37.
- Feuerstein, R., Rand, J., & Rynders, J.E. (1988). *Don't accept me as I am: Helping "retarded" people to excel*. New York: Plenum Press.
- Fletcher, J.M., Coulter, W.A., Reschly, D.J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia, 54*, 304-331.
- Foorman, B.R., Fletcher, J.M., Frances, D.J., Carlson, C.D. Chen, D., Mouzaki, A., et al. (1998). *Technical Report: Texas Primary Reading Inventory Technical*

(1998 Edition). Houston, TX: Center for Academic and Reading Skills and University of Houston.

Fuchs, D., Compton, D.L., Fuchs, L.S., Bryant, J., & Davis, G.N. (2008). Making “secondary intervention” work in a three-tiered responsiveness-to-intervention model: Findings from the first-grade longitudinal reading study of the national research center on learning disabilities. *Reading and Writing, 21*, 413-436.

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, 100*, 829-850.

Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice, 13*, 204-219.

Fuchs, D., & Fuchs, L. S. (2005). Responsiveness-to-intervention: A blueprint for practitioners, policymakers, and parents. *Teaching Exceptional Children, 38*, 57-61.

Fuchs, D., Fuchs, L.S., Benowitz, S., & Barringer, K. (1987). Norm-referenced tests: Are they valid for use with handicapped students? *Exceptional Children, 54*, 263-271.

Fuchs, D., Fuchs, L.S., & Compton, D.L. (2004). Identifying reading disabilities by responsiveness to instruction: Specifying measures and criteria. *Learning Disability Quarterly, 27*(4), 216-228.

- Gersten, R., & J.A. Dimino (2006). RTI (Response to Intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly*, 41, 99-108.
- Gillon, G. (2004). *Phonological Awareness: From research to practice*. New York: The Guilford Press.
- Good, R.H., Cummings, K.D., & Powell-Smith, K.A. (2008, February). ROC done right? Part 3 DIBELS benchmark goals. Paper presented at the annual Pacific Coast Research Conference, San Diego, CA.
- Good, R.H., & Kaminski, R.A. (2003). *Dynamic Indicators of Basic Early Literacy Skills*. Longmont, CO: Sopris West Educational Services.
- Good, R.H., Kaminski, R.A., Shinn, M., Bratten, J., Shinn, M., Laimon, L., et al. (2004). Technical adequacy and decision making utility of DIBELS Technical Report No. 7). Eugene, OR: University of Oregon.
- Gresham, F. K. (2002). Responsiveness to intervention: An alternative approach to the identification of Learning disabilities. In R. Bradley, L Danielson, D. Hallahan (Eds.) *Identification of Learning Disabilities: Research to Practice* (pp. 467-419). Erlbaum: Mahwah, NJ.
- Grigorenko, E.L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities*, 42, 111-132.
- Gutierrez-Clellen, V., Pena, E., & Quinn, R. (1995). Accommodating cultural differences in narrative style: A multicultural perspective. *Topics in Language Disorders*, 15, 54-67.

- Haager, D., Klingner, J., & Vaughn, S. (2007). *Evidence-based Reading Practices for Response to Intervention*. Baltimore, MD: Brookes Publishing Company.
- Hanley, J.A., & McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristics curves. *Radiology*, *148*, 839-843.
- Haywood, H.C., & Lidz, C.S.(2007). *Dynamic assessment in practice*. Cambridge: Cambridge University Press.
- Hogan, T.P., Catts, H.W., & Little, T.D. (2005). The relationship between phonological awareness and reading: Implications for the assessment of phonological awareness. *Language, Speech, and Hearing Services in the Schools*, *36*, 285-293.
- Individuals with Disabilities Education Improvement Act of 2004, Pub. L. 108-446, Stat. 2647.
- Janksy, J.J. (1978). A critical review of some developmental and predictive precursors of reading disabilities. In A.L. Benton & D. Pearl (Eds.), *Dyslexia: An appraisal of current knowledge* (pp. 313-347). New York: Oxford University Press.
- Jenkins, J. (2003). Candidate measures for screening at-risk students. Paper presented at the NRCLD Responsiveness-to-Intervention Symposium, Kansas City, MO.
- Jenkins, J.R., Hudson, R.F., & Johnson, E.S. (2007). Screening for at-risk readers in a Response to Intervention Framework. *The School Psychology Review*, *36*, 582-600.
- Jenkins, J.R., & Johnson, E.S. (2008). Universal screening for reading problems:

- When and how should we do this? Retrieved March 29, 2009, from
www.rtinetwork.org/Essential/Assessment/Universal/ar/ReadingProblems
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.
- Kingslake, B.J. (1983). The predictive (in)accuracy of on-entry to school screening procedures when used to anticipate learning difficulties. *British Journal of Special Education, 10*, 24-26.
- Kroese, J.M., Hynd, G. W., Knight, D.F., Hiemenz, J.R. and Hall, J. (2000). Clinical appraisal of spelling ability and its relationship to phonemic awareness (blending, segmenting, elision, and reversal), phonological memory, and reading in reading disabled, ADHD, and normal children. *Journal Reading and Writing, 13*, 105-131.
- Lewkowicz, N. K. (1980). Phonemic awareness training: What to teach and how to teach it. *Journal of Educational Psychology, 72*, 686-700.
- Lidz, C.S. (1987). *Dynamic Assessment: An Interactional Approach to Evaluating Learning Potential*. Guilford Press: New York.
- Lidz, C.S. (1991). *Practitioner's guide to dynamic assessment*. New York: Guilford.
- Lidz, C.S. (2005). Dynamic assessment with young children: We've come a long way baby! *Journal of Early Childhood and Infant Psychology, 1*, 99-112.
- Mann, V.A. (1993). Phoneme awareness and future reading ability. *Journal of Learning Disabilities, 26*, 259-269.
- Mann, V., & Wimmer, H. (2002). Phoneme awareness and pathways to literacy:

- A comparison of German and American children. *Reading and Writing: An Interdisciplinary Journal*, 15, 653-682.
- Marsh, G., & Mineo, R. J. (1977). Training preschool children to recognize phonemes in words. *Journal of Educational Psychology*, 69, 748-753.
- McBride-Chang, C. (1995). Phonological processing, speech perception and reading disability: An integrative review. *Educational Psychologist*, 30, 109-121.
- McBride-Chang, C., & Manis, F.R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. *Reading and Writing: An Interdisciplinary Journal*, 8, 323-339.
- Missiuna, C., & Samuels, M.T. (1988). Dynamic assessment: Review and critique. *Special Services in the Schools*, 5, 1-22.
- Muter, V., Hulme, C., Snowling, M.J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, 40, 665-681.
- National Center for Education Statistics (2007). The nations's report card: Reading 2007. Retrieved November 28, 2008 from <http://nces.ed.gov/nationsreportcard/>.
- National Reading Panel. (2000). *Report of the National Reading Panel*. Washington, DC: National Institute of Child Health and Development.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

- O'Connor, R.E., Fulmer, D., Harty, K., & Bell, K. (2005). Layers of Reading Intervention in Kindergarten through Third Grade: Changes in Teaching and Child Outcomes. *Journal of Learning Disabilities, 38*, 440-455.
- O'Connor, R.E., & Jenkins, J.R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3*, 159-197.
- Pena, E.D., Gillam, R.B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research, 49*, 1037-1057.
- Pennington, B.F., & Lefly, D.L. (2001). Early reading development in children at family risk for dyslexia. *Child Development, 72*, 816-833.
- Perfetti, C.A., Beck, I., Bell, L.C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283-319.
- Rathvon, N. (2004). *Early reading assessment: A practitioner's handbook*. New York: Guilford Press.
- Sackett, D.L., Haynes, R.B., Guyatt, G.H., & Tugwell, P. (1991). *Clinical epidemiology: A basic science for clinical medicine*. Boston: Little, Brown.
- Salvia, J., & Ysseldyke, J. (1988). *Assessment in special and remedial education* (4th ed.). Boston, MA: Houghton-Mifflin.
- Savage, R.S., & Fredrickson, N. (2005). Further evidence of a link between rapid naming and text reading rate. *Brain and Language, 93*, 152-159.

- Scanlon, D.M. & Vellutino, F.R. (1996) Prerequisite skills, early instruction, and success in first-grade reading: Selected results from a longitudinal study. *Mental Retardation and Development Disabilities Research Reviews*, 2, 54-63.
- Scanlon, D.M., Vellutino, F.R., Small, S.G., Fanuele, D.P., & Sweeney, J.M. (2005). Severe reading difficulties--Can they be prevented? A comparison of prevention and intervention approaches. *Exceptionality*, 13, 209-227.
- Scarborough, H.S. (1998). Early identification of children at risk of reading difficulties: Phonological awareness and some other promising predictors. In B.K. Shapiro, P.J. Accardo, & A.J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75-120). Timonium, MD: York.
- Schatschneider, C., Fletcher, J.M., Francis, D.J., Carlson, C.D., & Foorman, B.R. (2004). Kindergarten Prediction of Reading Skills: A Longitudinal Comparative Analysis. *Journal of Educational Psychology*, 96, 265-282.
- Schatschneider, C., Francis, D.J., Foorman, B.R., Fletcher, J.M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology*, 91, 439-449.
- Siegel, L.S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities*, 25, 618-629.
- Simmons, D., Coyne, M., Kwok, O., McDonagh, S, Harn, B., & Kame'enui, E. (2008). Indexing response to intervention: A longitudinal study of reading risk

- from kindergarten through third grade. *Journal of Learning Disabilities*, 41, 143-158.
- Spector, J.E. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology*, 84, 353-363.
- SPSS for Windows, Rel. 16.0.1. 2007. Chicago: SPSS Inc.
- Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, 86, 221-234.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-407.
- Stanovich, K. E., Cunningham, A. E., & Cramer, B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology*, 38, 175-190.
- Sternberg, R.J., & Grigorenko, E.L. (1998). *Dynamic Assessment: The Nature and Measurement of Learning Potential*. Cambridge: Cambridge University Press.
- Swets, J.A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*, 47, 522-532.
- Swets, J.A., Dawes, R.M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1-26.
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.
- Torgesen, J.K. (1998). Catch them before they fall. *American Educator*, 32-39.

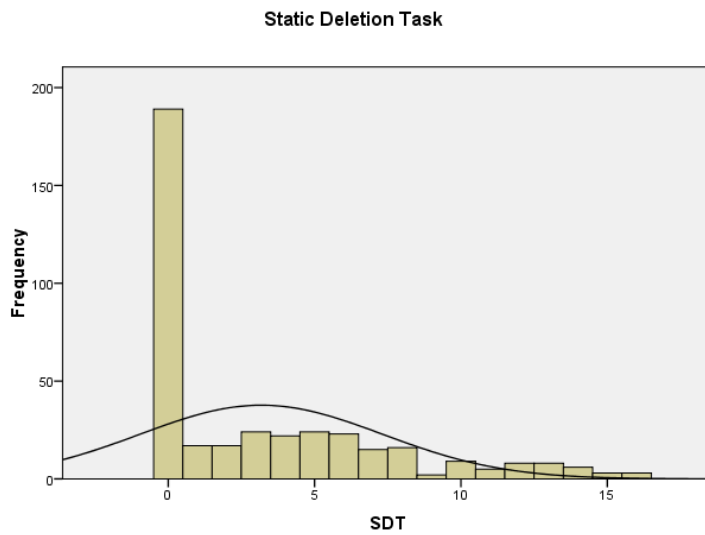
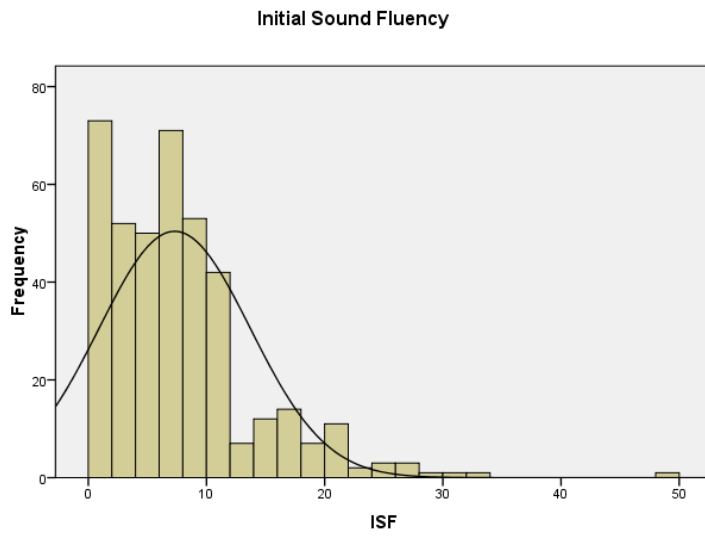
- Torgesen, J.K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 15*, 55-64.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). The development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology, 30*, 468-479.
- Tzuriel, D., & Klein, P. (1987). Assessing the young child: Children's analogical thinking modifiability. In C.S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 268-287). New York: Guilford.
- VanDerHeyden, A., Witt, J.C., & Gilbertson, D. (2007). A multi-year evaluation of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45*, 225-256.
- Vaughn, S., & Fuchs, L.S. (2003). Redefining learning disabilities as inadequate response to intervention: The promise and potential problems. *Learning Disabilities Research and Practice, 18*, 137-146.
- Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33*, 321-363.
- Vellutino, F.R., Scanlon, D.M., Sipay, E.R., Small, S.G., Pratt, A., Chen, R.S., et al.

- (1996). Cognitive profiles of difficult to remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, 88, 601-638.
- Vellutino, F.R., Scanlon, D.M., Small, S., & Fanuele, D.P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities*, 38, 157-169.
- Vellutino, F.R., Scanlon, D.M., & Zhang, H. (2008). Identifying reading disability based on response to intervention: Evidence from early intervention research. In S.R. Jimerson, M.K. Burns, & A.M. VanDerheyden (Eds.) *The Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention*.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MS: Harvard University Press.
- Wagner, R., & Sternberg, R.J. (1984). Alternative conceptions of intelligence and their implications for education. *Review of Educational Research*, 54, 179-223.
- Wagner, R.K., & Torgesen, J.K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 30, 73-87.

- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: PRO-ED.
- Whitehurst, G.J., & Lonigan, C.J. (1998). Child development and emergent literacy. *Child Development, 69*, 848-872.
- Woodcock, R.W. (1998). *Woodcock Reading Mastery Tests-Revised Normative Update*. Circle Pines, MN: American Guidance Service.

FIGURE 2.1

Histograms associated with the Initial Sound Fluency measure, the Static Deletion Task, and the Dynamic Screening of Phonological Awareness from Study 1



Dynamic Screening of Phonological Awareness

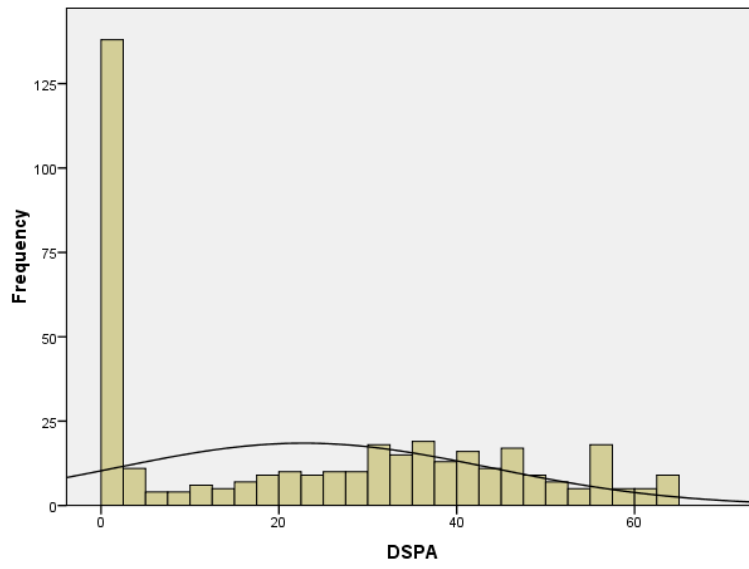
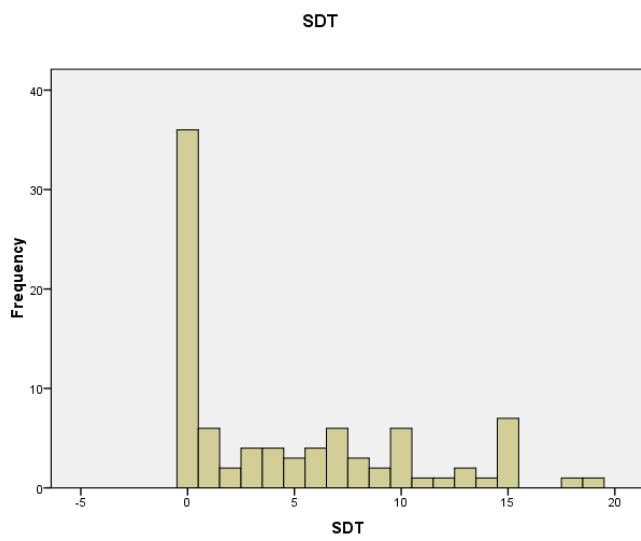
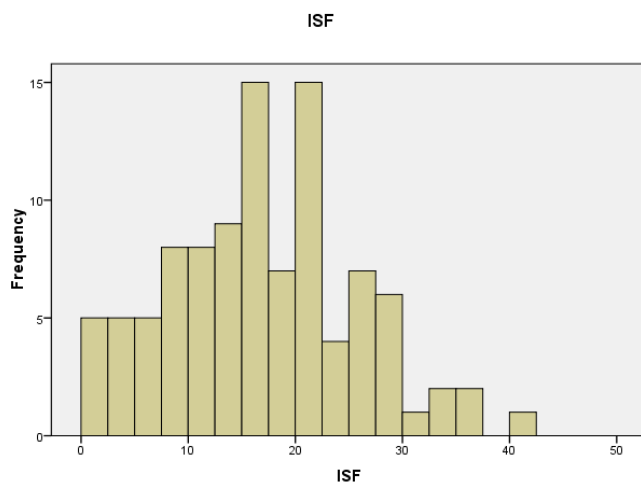


Figure 3.1
Histograms for Initial Sound Fluency, Static Deletion Task, and Dynamic Screening of Phonological Awareness-Revised for Sample 1 (N=90)



DSPA-R

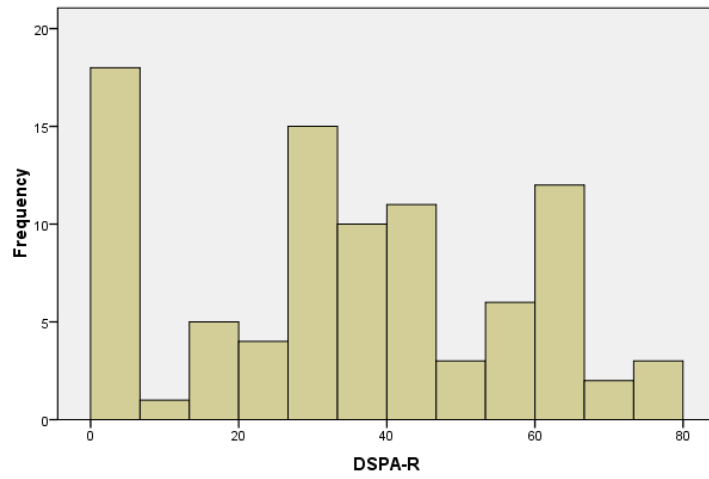


Figure 3.2
Histograms for the Initial Sound Fluency measure and the Dynamic Screening of Phonological Awareness-Revised for the full Sample 2 (N=161)

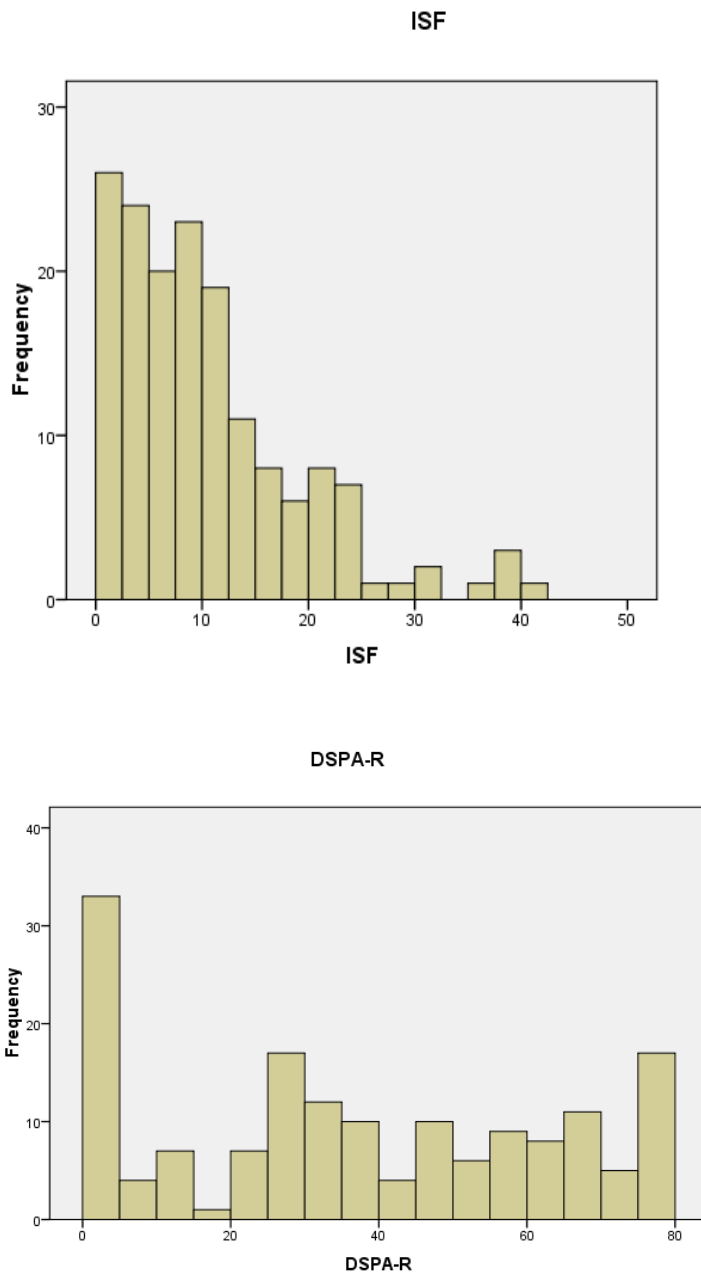
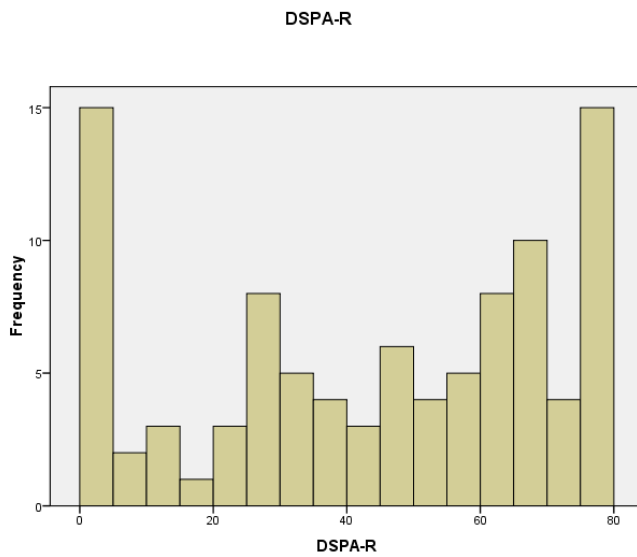
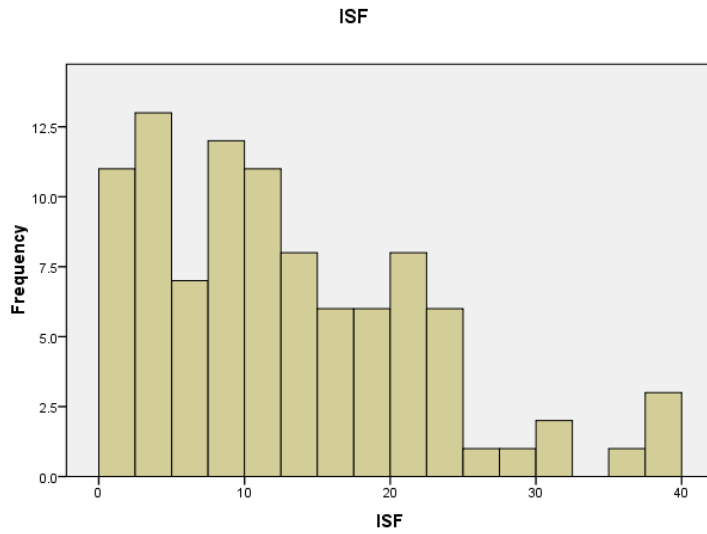


Figure 3.3
Histograms for the Initial Sound Fluency measure and the Dynamic Screening of Phonological Awareness-Revised for the reduced Sample 2 (N=96)



Appendix A

List of items for DSPA (Study I)

Initial syllable deletion (Items 1-8)

1. doghouse
2. inside
3. pretest
4. mildew
5. tulip
6. motel
7. dolphin
8. pony

Initial phoneme deletion (Items 9-16)

9. sit
10. fan
11. cough
12. shout
13. make
14. twin
15. plate
16. snail

Appendix B

DSPA Prompt Protocol

PROMPTS for items 1-8 (syllable deletion):

Prompt 1: Repeat the initial example sentence and then repeat the stimulus question.

“Remember, the word ‘sailboat’ without sail is boat. Now, say (**doghouse**) without (**dog**).”

Prompt 2: Pause after the initial syllable and then emphasize the remaining portion of the word.

“Try again. Say (**doghouse**) without (**dog**).”

Prompt 3: Tap the square on your right while you say the first syllable, pause, then tap the square on your left as you say the remaining portion of the word with emphasis.

“Try again. Say (**doghouse**) without (**dog**).”

Final answer: Give the answer in scripted form.

“Doghouse without dog is house.”

PROMPTS for items 9-16 (sound deletion):

Prompt 1: Repeat the initial example sentence and then repeat the stimulus question.

“Remember, the word shin without “sh” is in. Now, you say ____ without ____.”

Prompt 2: Pause after the initial sound and then emphasize the remaining portion of the word.

“Try again. Say (**sit**) without (**/s/**).”

Prompt 3: Tap the square on your right while you say the first sound, pause, then tap the square on your left as you say the remaining portion of the word with emphasis.

“Try again. Say (**sit**) without (**/s/**).”

Final answer: Give the answer in scripted form.

“(**Sit**) without (**/s/**)”is (**it**).”

Appendix C: Fidelity checklist

DSPA
Fidelity Checklist

Examiner:

Time and Date:

+	-
	<p>Examiner places blocks (for visual feedback) in front of child prior to starting DAPA. Examiner delivers initial instructions verbatim.</p> <p>Examiner provides initial model. Examiner pronounces every item correctly.</p> <p>Examiner administers every item in the correct order. Examiner administers the first prompts correctly (when appropriate).</p> <p>Examiner uses appropriate stress cue the second prompts correctly (when appropriate). Examiner uses appropriate pausing on the second prompts correctly (when appropriate).</p> <p>Examiner uses appropriate visual cue on the third prompts (when appropriate).</p> <p>Examiner provides correct target if child does not answer correctly after all prompts are given. Examiner gives child ten seconds of pause time after asking question or administering prompt. Examiner continues or discontinues after five consecutive zero scores.</p>

Appendix D

List of items for DSPA-R (Study II)

Syllable deletion

1. doghouse
2. football
3. pancake
4. rainbow
5. midnight
6. untie
7. pretest
8. repair
9. tulip
10. dolphin
11. pony

Phoneme deletion

12. fan
13. sit
14. cough
15. shout
16. make
17. twin
18. snail
19. plug
20. crave