PREDICTING STUDENT RESPONSIVENESS TO FAST FORWORD USING DIBELS SUBTESTS

A Dissertation

Submitted to

the Temple University Graduate Board

In Partial Fulfillment

of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

By

Fernando Cavallo

January, 2012

Examining Committee Members:

Catherine Fiorello, Advisory Chair, School Psychology
Joseph Rosenfeld, School Psychology
Jospeh DuCette, Educational Psychology
Frank Farley, School Psychology
Kenneth Thurman, Curriculum, Instruction, and Technology in Education

ABSTRACT

Predicting Student Responsiveness to Fast ForWord Using DIBELS subtests
Fernando Cavallo
Doctor of Philosophy
Temple University, 2012
Doctoral Advisory Committee Chair: Catherine A. Fiorello, Ph.D.

The current study was completed through a retrospective analysis of school records of elementary school students in the Northeast Region of the Philadelphia School District (PSD) who have participated in the Fast ForWord (FFW) Language program. The data requested from student records included: demographic information (e.g., gender, grade,age, ethnicity, disabilty, and special education status), DIBELS scores, and FFW completion and participation variables.

The current study set out to determine if DIBELS scores can predict student performance or mastery level on the FFW program. A total of seven individual FFW variables (percent of completion for each activity) and five overall FFW variables (percent complete, participation level, attendance level, total days to complete, and successful performance) served as the outcome variables. Frequency distributions, Pearson correlations, an ANOVA, and a standard multiple regression were used to determine the relationships of demographic variables among predictor and outcome variables as well as the predictive power of the DIBELS test scores.

Results of the standard multiple regression analysis failed to yield significant results in the ability for either DIBELS raw or benchmark scores to predict performance on the FFW reading program. The current study highlighted that in the real-life conditions of a large, urban public school system, DIBELS seems to have very little, if any, predictive abilities specific to designating students appropriately

to an intensive, costly, and time-consuming intervention program. It is hoped that the information presented in this study will stimulate some positive discussion and changes in the initial assessment and referral processes currently being widely employed across American schools in order to better serve and educate American children that demonstrate symptoms of early reading deficits.

ACKNOWLEDGEMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my family and wife. The printed pages of this dissertation represent far more than the culmination of years of study. These pages also reflect the relationships with many generous and inspiring people I have met along the way, many of which have become dear friends, colleagues, and even a spouse.

I would like to express my sincerest gratitude to my advisor, Dr. Catherine Fiorello, for her excellent guidance and patience. Her confidence in me served as a neverending source of inspiration for me to persevere and overcome the numerous hurdles I encountered. I simply could not have completed this dissertation without her expertise and gracious supervison.

To my committee members Drs. Joseph DuCette, Frank Farley, Joseph Rosenfeld, and Kenneth Thurman for their encouraging words, thoughtful criticism, and time and attention during busy semesters. I would also like to thank the many professors who instructed me during my graduate career for showing me by example and through challenging coursework how to critically analyze information and strive to be a consummate professional.

I am also grateful to the School District of Philadelphia for supporting my research agenda as the present research study would have not been possible without their cooperation. Special thanks to Faith Sproul, a fellow graduate student and friend who generously and repeatedly assisted me during the process.

I would also like to express a personal thanks to all of those people, friends and colleagues, who have touched my life in a special way and continue to guide me in my professional and personal career throughout life. While the list is long, I would like to specifically recognize my professional mentor, Carole Gordon-Young for providing me not only with an examplary model of how to be a great school psychologist, but one of my biggest supporters in my educational pursuits. Special thanks also to Dr. Joe Johnston, who is not only a colleague in both research and work, but a good friend in life who was always willing to help and give me suggestions even when I did not solicit them.

I would also like to thank my parents, Fernando and Lidia Cavallo, and my two brothers, Rocco and Anthony. They were always supporting and encouraging me to pursue my professional and academic goals even when the road was long and foggy.

Finally, I would like to express my deepest gratitude to my wife, Rossella, who selflessly encouraged me to return to graduate school to complete my doctoral degree. She never waivered in her support despite the countless sacrifices she made during what seemed to be a neverending pursuit at the time. Rossella has always been there graciously encouraging me through the good times and bad and always refocused me when I would seem to get lost along the way. Perhaps the greatest creation of all during this process is our precious little daughter, Julia, who inspires me to be the best father and person I can be. Her smile and energy intoxicates me with joy every day.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Reading Interventions in Today's Educational Context

Reading plays a central role in education, and the first few years of school typically focus on teaching children to read. However, learning to read is not a simple process. Rather, reading is a complex process that involves many components and cognitive skills. Many definitions exist for the term 'reading,' and there are different components to reading, such as reading rate, accuracy, and fluency. The cognitive skills involved in reading differ at various stages of learning to read (Riedel, 2007; Robeck & Wallace, 1990). Difficulty acquiring reading skills is the most common referral reason for educational assessment (Ramus, 2001).

Psychological and educational investigators have attempted to develop models of reading that identify the cognitive skills necessary for learning to read, reading well, and reading efficiently. Before a child can begin learning to read, a child must understand that printed letters stand for units of speech. An inability to name letters in young children has been found to predict later failures at reading. An ability to discriminate between the sound parts of words, or auditory discrimination, is also believed to be a basic skill for learning to read (Robeck & Wallace, 1990). The list for skills necessary for reading is long, and includes phonemic segmentation, phonemic discrimination, visual-verbal association, visual perception and discrimination, visual memory, auditory memory, intelligence, language comprehension, sequencing, and visual and auditory attention (Downing & Kan Leong, 1982; Robeck & Wallace, 1990; Vanderwood et al., 2002).

As reading is such an integral part of learning in school, many interventions have been developed that attempt to improve the reading skills of children. Fast ForWord (FFW) is a relatively new, computer-based intervention developed based on neuroscientific research. The FFW program aims to improve the cognitive and language skills in children that are necessary for reading (Scientific Learning Corporation, 2006a). The program utilizes acoustically modified speech because of research demonstrating that children with language and reading impairments have temporal processing deficits. A temporal processing deficit refers to a deficient ability to process brief acoustic stimuli that occur in rapid succession (Tallal, 2000) and has been found to be correlated with specific reading disorders such as dyslexia (Shaywitz, 1998). Neuroscientific research suggests that phonemes of a language must be learned from exposure to the environment. These phonemes may be represented in the brain as distinct neuronal firing patterns (Tallal, 2000). Research on neuroplasticity suggests that it may be possible to re-train how the brain processes rapidly occurring sensory stimuli, such as phonemes (Tallal, 2000).

In the FFW intervention, there are different activities and levels that children complete. The different activities target various language and reading skills, such as listening accuracy and phonological awareness. As students progress through the different levels of each activity, the acoustically modified sounds and speech gradually approximate normal language (Learning Corporation, 2006a). There are many different FFW products, including FFW-L and FFW Language to Reading

(FFW-LR). These programs are currently being used in the Philadelphia School District as a reading intervention.

Conceptual Framework for Response to FFW Intervention

For the special education population and for those students suspected of having a learning disability, improving reading is often a significant and primary goal. The Scientific Learning Corporation claims that the FFW programs benefit a wide variety of students who have reading difficulties. However, the few research studies published often excluded students with below normal intelligence (e.g., Cohen, Hodson, & O'Hare et al., 2005; Hook, Macusaro, & Jones, 2001). At the same time, other research studies indicate that FFW may be more effective for students who are the most impaired (e.g., Troia, 2004). More data are needed to determine if FFW is an effective reading intervention, particularly with special education students.

The FFW programs are being implemented in many school districts across the United States, including the Philadelphia School District. Low standardized test performance and high drop out rates are troubling problems in the Philadelphia School District. In order to improve reading scores, the Philadelphia School District has invested in the FFW programs. Computer availability, teacher schedules, and length of the programs limit the number of students who can participate in the FFW programs. In the Philadelphia School District, at the time of this study, there were approximately 170,000 total students and 23,596 special education students. Of these students, 4.4%, or 7,483 total students, participated in FFW programs during the 2007-2008 school year. In the Northeast Region alone there are approximately 27,395 total students and 4,035 special education students. In the Northeast Region, 39% of students (approximately 10,684 students) do not meet target reading levels. It is apparent that the population of students who could potentially benefit from participation in FFW far exceeds the number of spots available for participation.

Numerous studies have demonstrated the efficacy of the FFW programs for improving language and reading. As with any intervention, the literature also implies that some students may not respond as well to the FFW program as others (Deppeler, Taranto, & Bench, 2004; Scientific Learning Corporation, 2004; Troia, 2004; Troia & Whitney, 2003; Hook, Macaruso, & Jones, 2001; Loeb, Stoke, & Fey, 2001). Therefore, when selecting students to participate in FFW, it would be beneficial to know in advance which students are most likely to respond well to the program and, therefore, would be more likely to experience improved language and reading skills. This is especially important considering the average length of time it takes to complete the FFW interventions (approximately 6-8 weeks of daily sessions per program), limited teacher availability and scheduling concerns, and limited computer availability.

Purpose of the study

The purpose of the proposed study was to determine if literacy screeners typically given in the school setting (Dynamic Indicators of Early Literacy Skills-DIBELS) can help identify students who will benefit from and achieve mastery on the FFW-Language (FFW-L) program.

The primary research questions addressed in this study were:

1. Can any of the DIBELS subtests significantly predict which students will achieve the mastery level on the FFW-L program?
2. Which specific DIBELS variables significantly predict those students that will benefit from the FFW-L program?
3. Do demographic variables predict which students will successfully complete the FFW-L program?

Definition of Key Terms

For clarity, the following definition of terms will be used in this study.

Scientific Learning Corporation (SLC): software company (registered trademark) which developed the Fast Forward product.

FFW Language: (Fast ForWord)- computer-based software program designed by the Scientific Learning Corporation designed to improve literacy and reading skills.

FFW rules of completion: refers to the criteria established by SLC in order to determine if students have meaningfully and successfully completed the FFW training programs. FFW has defined successful completion to have been met if the student achieves a 90% total completion score or 80% complete on a majority of both sound and word exercises with notable and steady progress.

FFW attendance level: refers to whether a student is following the prescribed protocol of using the program daily for 50 minutes a day, five days a week, for a consecutive 6-16 weeks and translates this score into a percentage. It essentially measures if the student is using the program consistently. For example, if a student works on the program 4 out of 5 days, his attendance level would be 80%.

FFW percent complete: represents the percentage of the product the participant has successfully covered. The score helps illustrate the participant's understanding of the concepts covered in a particular product and its cumulative. FFW's established goal is for a participant to achieve 2-5% progress toward completion daily, so after 30 sessions, the student should have a minimum of 60% of the program completed.

FFW participation level: measures a student's time on-task on the days they show up to work on FFW The percent of time the student stayed logged-on and working. If a student logs off or stops working after 40 minutes on a 50-minute protocol, then their participation level = 80%.

DIBELS: (Dynamic Indicators of Basic Early Literacy Skills)- a literacy measures which assess the five main areas in early literacy identified by the National Reading Panel: phonemic awareness, alphabetic principle, accuracy and fluency, vocabulary, and comprehension.

Positive Responders: students who do respond favorably or demonstrate significant benefits from the FFW Language program according to completion criteria established by Scientific Learning Corporation.

Non-Responders: students who do not respond favorably or fail to demonstrate significant benefits from the FFW Language program according to completion criteria established by Scientific Learning Corporation.

Special Education students: students receiving IEP services due to a qualifying handicapping condition in accordance with the Pennsylvania Department of Education special education regulations.

Response to Intervention (RTI): a new model enacted in IDEA 2004 as an alternative to the discrepancy model for diagnosing learning disabilities. RTI also introduced a national three-tiered system for monitoring students' academic and behavioral progress.

IDEA: The Individuals with Disabilities Education Act (IDEA) is a United States federal law hat governs how states and public agencies provide early intervention, special education, and related services to children with disabilities. It addresses the educational needs of children with disabilities from birth to age 21.

AIMSweb: AIMSweb is a benchmark and progress monitoring system based on direct, frequent and continuous student assessment. The results are reported to students, parents, teachers and administrators via a web-based data management and reporting system to determine response to intervention.

Corrective Reading: Corrective Reading is designed to promote reading accuracy (decoding), fluency, and comprehension skills of students in third grade or higher who are reading below their grade level. The program has four levels that address students' decoding skills and six levels that address students' comprehension skills. All lessons in the program are sequenced and scripted. Corrective Reading can be implemented in small groups of four to five students or in a whole-class format.

Corrective Reading is intended to be taught in 45-minute lessons four to five times a week.

Reading First Funds: Reading First is a nationwide effort providing funds to local education agencies and public or private organizations that serve children from low-income families. Federal funds are awarded competitively to local programs that show they will enhance young children's language and cognitive development by providing high-quality instruction and ongoing professional development based on scientifically based research.

CHAPTER 2

REVIEW OF LITERATURE

The RTI model for Educational Intervention

The reauthorization of the Individual with Disabilities Education Act (IDEA) 2004 (Federal Register, 2006) has clearly challenged and scrutinized the legitimacy of diagnosing learning disabilities through the traditional discrepancy model approach. In an attempt to remediate the over-identification of specific learning disabilities which has occurred in the past 30 years, federal legislation (IDEA, 2004) has moved toward a response to intervention or RTI model which claims to utilize a more rigorous and scientifically-based approach to identifying the presence of a specific learning disability. Proponents of this new model (Burns et al., 2006; Kovaleski, 2006; VanDerHeyden, Witt, & Nacquin et al., 2003) discourage the use of intellectual assessments, such as standardized IQ measures, in the identification process of a specific learning disability (SLD) due to over-identification which occurs when school psychologists employ the discrepancy model to compare children's intellectual functioning or potential with their actual level of academic achievement at the time of the assessment. Instead, proponents of the RTI model and IDEA 2004 suggest that the American educational system should shift to an SLD identification model, namely RTI, which examines whether a child has responded to "scientifically, research-based intervention."

RTI proponents suggest that we use an SLD identification process based on systematic assessment of the student's response to high quality, research-based general education instruction. Such a process essentially implements a hierarchical series of research-based interventions to students at-risk (or thought to be eligible) of a specific learning disability. If the student "fails to respond" or fails to make "appropriate" academic progress in the allotted intervention period, which is stipulated to typically consist of four to six weeks of intensive research-based intervention, then the student could be identified with an SLD and determined to be eligible for and in need of specially designed instruction.

Based on the RTI model, there are three tiers or levels of intervention services provided to students deemed to be at-risk for significant academic deficits. Tier 1 in the RTI model is typically implemented for a group of students in a given classroom who stand-out as being "at-risk" due to their low performance in a given area (i.e., reading) according to a universal screening measure. Therefore a Tier 1 instruction plan might consist of a guided reading group conducted daily in the regular classroom setting for students who are performing at the lowest reading level in the classroom. Students' progress during each of the tiered interventions is monitored in order to assess their response to the actual intervention. The DIBELS is a tool commonly used by school districts to assess and identify students' reading skills. Students are typically recommended for Tier 1 interventions based on DIBELS test scores. In the RTI model, assessment is used for the purposes of screening, collecting diagnostic information, and monitoring progress. The RTI model does specify that the students' level of need dictates the level of tier services they receive (Burns et al., 2006). Therefore, a student who displays severe academic needs may skip Tier 1 and move directly to Tier 2 for more intensive and individualized intervention. Typically, students advance from Tier 1 to Tier 2 when they fail to respond to Tier 1 interventions, thus requiring additional academic supports including the manipulation of instructional time and instructional focus beyond what was provided in Tier 1. A Tier 2 intervention could include a referral to an intensive reading intervention, such as the FFW-L training program.

The actual length of time that an intervention is implemented depends on the student's response to the intervention and realistic time periods required for the target skills to develop. It is also possible that students will receive interventions in more than one area and might actually receive a different level of tier interventions simultaneously due the fluid nature of the RTI continuum of tiered services (Burns et al., 2006). For example, a student might receive Tier 1 intervention for math related weaknesses and Tier 2 services for identified reading weaknesses. Students who improve as a result of Tier 2 interventions may be moved back to a Tier 1 level of intervention, which includes interventions accessible in the general classroom setting. Students who are not successful with Tier 2 services despite appropriate interventions over time may be considered for a Tier 3 level of services, which involve long-term interventions and may lead to the provision of special education services.

Progress Monitoring

The critical component underlying the RTI process involves the collection of progress monitoring data necessary to make formative and summative evaluations of the students' response to intervention. Jenkins and Johnson (2009) suggest that universal screening procedures should be quick and easy to administer and interpret. However, Jenkins and Johnson (2009) also note that research has demonstrated that a screening battery which uses multiple measures provides better classification accuracy for distinguishing at-risk and typically developing students than any single-measure screening approach. Despite this, many school districts across the nation rely on single-measure assessments, such as the DIBELS, state-based benchmark tests, or proficiency tests to serve as both their initial screening device and progress monitoring assessment tool. Perhaps this is due to the fact that these assessments occur routinely throughout the school year and are "built-in" to the school calendar and curriculum. Although the RTI model does not specify which screening and progress monitoring procedures are deemed appropriate or inappropriate for the implementation of RTI, certain progress monitoring tools have received positive endorsements from national organizations such as the National Assessment of Educational Progress (NAEP), National Association of School Psychologists (NASP), etc. AIMSweb is perhaps the most popular RTI assessment and data management system publicized and supported by national educational organizations. In fact, AIMSweb was awarded the 2009 Best in Tech Award by the "Scholastic Adminmistrator." In addition, the National Center on RTI endorsed AIMSweb screening and progress monitoring assessments with their highest rating for reliability and validity. According to the AIMSweb website (n.d.), "AIMSweb is a benchmark and progress monitoring system based on direct, frequent and continuous student assessment. The results are reported to students, parents, teachers and administrators via a web-based data management and reporting system to determine response to intervention." Similar to the DIBELS, AIMSweb provides a benchmark of all students three times a year: Fall, Winter, and Spring. However, AIMSweb provides benchmarking as well as progress monitoring assessment procedures in the following areas: early literacy, reading, language arts, and math. Progress monitoring data for AIMSweb rely on a curriculum-based measurement methodology. AIMSweb developers proudly advertise that their progress monitoring assessment measures are "quick and easy" to administer (1 to 4 minutes) and score (n.d.). Furthermore, progress monitoring assessment can be administered monthly as an individual or group assessment. Currently, AIMSweb and DIBELS appear to be the most common tools utilized nationally in schools for benchmark and progress monitoring assessment. However, since the RTI model and IDEA 2004 (Federal Register, 2006) do not specify specific criteria or standards for progress monitoring instruments, individual school districts may employ their own progress monitoring measures, such as curriculum-based assessments or state benchmark tests. Finally, new RTI progress monitoring assessment packages other than AIMSweb are being developed and have already influenced traditional achievement test developers to incorporate a progress monitoring system into individually administered achievement tests which are typically included in a psychoeducational battery. For example, the Wide Range Achievement Test- Fourth Edition (WRAT-4) developed by Wilkinson and Robinson (2006) has a progress monitoring version of the test: WRAT4-PMV.

DIBELS

DIBELS is widely used in the United States and has gained widespread popularity in American schools following the NCLB Act (2001). Tierney and Thome (2006) credit the NCLB Act as the impetus to large scale reading initiatives in this country. According to Samuels (2007), an estimated 1,800,000 students are assessed yearly using the DIBELS. However, some researchers such as Riedel (2007) question whether its popularity is due more to political influences rather than research-based support. In fact, Tierney and Thome express their concern that it is likely not a coincidence that the five essential components for reading established by the National Reading Panel in 2000 match the same five skills purported earlier to be measured by the DIBELS.

According to the creators of DIBELS (Good & Kaminski, 2002) it was designed to be a brief, economical, and efficient indicator of a student's progress toward achieving a general outcome in the acquisition of early literacy skills from kindergarten through sixth grade. Brunsman (2005) indicates that the DIBELS provides a screening of reading skills and is considered primarily useful for tracking students' reading progress in the early grades. The authors of the DIBELS claim that it is useful in predicting future reading difficulties and identifying students in need of intervention. Brunsman (2005) points out that DIBELS is primarily effective in: 1) measuring critical skills that underlie reading success, 2) predicting success on high-stakes tests such as the Pennsylvania System of School Asssessment (PSSA), and 3) evaluating the effectiveness of reading instruction. However, Brunsman

(2005) warns that the DIBELS was not meant to be utilized as the sole diagnostic measure for identifying reading disabilities. Various reviewers of the DIBELS (Brunsman, 2005; Nelson, 2008; Riedel, 2007) suggest the following advantages of DIBELS: 1) it is cheap and easy to administer, 2) it is practical and effective for screening and monitoring students' reading progress and, 3) the administration and scoring guide has detailed information. In contrast, these same reviewers provide a litany of disadvantages associated with the use of the DIBELS.

The primary disadvantages of the DIBELS focus on the test's lack of statistical power. That is, Brunsman (2005) points out that the DIBELS manual contains insufficient reliability and validity measures and failed to establish any degree of predictive validity. Furthermore, Brunsman (2005) noted that research evidence does not exist to support the scope of the DIBELS as purported by the test developers. In addition, the instructional categories established by the test creators (at-risk, some-risk, low risk) are not reliable according to Brunsman (2005) and lacks research evidence to support the creation of these categories. In fact, the "at-risk" category is defined by performance falling below the 20^{th} percentile and "some-risk" is defined by performance falling between the 20^{th} – 40^{th} percentile according to localized norms. Statistically, half of the population taking the DIBELS would, by statistical definition, perform in the 25^{th} – 75^{th} percentiles. Therefore, cut-off scores established by the DIBELS appear to be overly broad which likely leads to a significant amount of false positives when identifying students in need of remedial reading interventions. In fact, Brunsman (2005) criticized the DIBELS for its inability to provide or gain evidence-based support for the use of the instructional based categories in order to effectively recommend intervention for specific student groups. According to Brunsman (2005) the predictive power of the DIBELS has not been proven and is meager at best.

Nelson (2008) conducted a correlational study in order to investigate the classification validity of the DIBELS. Nelson administered both the DIBELS and the Test of Phonological Awareness (TOPA-2+) to kindergarten students in January. He then administered the letter word identification and word attack subtests of the Woodcock Johnson Achievement Battery- 3rd Edition (WJ-III) four months later. The results revealed that 24.9% of the sample scored below the TOPA-2+ cutoff, whereas 10.2% of the sample scored below the WJ-III cutoff. These results led Nelson to conclude that the DIBELS cut-off score for defining the “at-risk” group had low sensitivity when the WJ-III subtests were used as a measure of future reading skills. Thus, Nelson also concluded that the cut-off scores for the “at-risk” and “some-risk” criteria are not reliable and strongly suggested that these criteria need to be adjusted. In fact, Nelson found that the DIBELS missed 68% of the sample for at-risk readers when used as the sole screening measure. Interestingly, in his study Nelson found that the DIBELS was better at negative predictive power versus positive predictive power. That is, the DIBELS was better at predicting and identifying good readers, but poor at identifying inadequate readers. Although one could argue that over identifying false positives is not necessarily harmful to students based on the notion that extra resources are provided when not indicated, Bishop (2003) warns that false positives on such high-stakes tests squander meager instructional resources as well as promote unnecessary parent and teacher concern as well as student anxiety. Based on Nelson's findings, the DIBELS can be generally criticized for not being practical and over identifying and prescribing students for intervention. This has caused critics to question whether the significant investment of time and money is justified or simply a waste of school-based resources.

Another study by Riedel (2007) examined the relationship between DIBELS, reading comprehension, and vocabulary skills among first grade students in an urban school setting. Riedel sought to further explore the widely held criticism that DIBELS does not truly measure reading comprehension, but merely a student's ability to read quickly as assessed by the Oral Reading Fluency (ORF) subtest on the DIBELS . Riedel suggested that reading fluency involves both speed and comprehension skills and that the ultimate goal of developing basic reading skills is to develop adequate reading comprehension skills. Therefore, Riedel conducted a retrospective research study in which he correlated DIBELS scores administered on three separate occasions during the course of the year to first grade students with reading comprehension scores from the Group Reading Assessment and Diagnostic Evaluation (GRA+DE) test administered at the end of first grade and TerraNova reading test scores administered in the second grade. Riedel did not include any special education students in his sample. The results revealed that the ORF subtest of the DIBELS was a good predictor of reading comprehension at the end of first and second grade with an 80% and 71% accuracy, respectively. In addition, the remaining DIBELS subtests were less accurate in predicting reading comprehension with Phonemic Sound Fluency (PSF) identified as the weakest predictor. Riedel also demonstrated that ORF combined with additional DIBELS subtests did not improve the predictive power of ORF alone with regard to performance on the GRA+DE and TerraNova tests. This led Riedel to conclude that for first grade students, the ORF subtest of the DIBELS is enough to predict reading comprehension skills and that administration of the other DIBELS subtests is unnecessary.

Additionally, critics such as Samuels (2007) argue that the DIBELS subtests do not accurately measure reading processes in early elementary grade students. In fact, Samuels asserts that the creators of DIBELS wrongly include the word "fluency" in the subtest names. Samuels suggests that the nature of the DIBELS subtests ignores the notion that beginning and emergent readers engage in two different reading processes. More specifically, Samuels explains that beginning readers focus their cognitive resources on decoding skills, not comprehension; whereas the emergent/fluent reader simultaneously engages in decoding and comprehension skills during reading tasks. Instead of recognizing these two distinct processes, Samuels notes that the DIBELS subtest instructions focus exclusively on speed and argues that the task instructions do not tell students to read for comprehension. In fact, Samuels also noted that advanced readers tend to comprehend less if they read too fast, yet DIBELS instructs students to read as fast as possible. Perhaps this could explain why the DIBELS has been repeatedly found to significantly over identify false positives.

Hintze et al. (2003) also assessed the concurrent validity and diagnostic accuracy of the DIBELS by investigating whether the DIBELS was predictive of student performance on the Comprehensive Test of Phonological Processing

(CTOPP) using both the DIBELS original criterion scores as well as alternative cut-off scores suggested by Hintze, et al. The CTOPP test developers, Torgesen and Rashotte (1999), report the CTOPP to have moderate to strong criterion prediction validity with the Gray Oral Reading Test-Third Edition (GORT-3), Woodcock Reading Mastery Test-Revised(WRMT-R), Wide Range Achievement Test-Third Edition (WRAT-3), and Test of Word Reading Efficiency (TOWRE). Inter-rater reliability ranged from .95 to .99 for all ages. Despite the CTOPP's technical superiority over the DIBELS as a measure of phonological reading skills, Torgesen and Rashotte note that the CTOPP manual suggests that additional assessment procedures be used in conjunction with the CTOPP in order to design instructional intervention for individual students. The experimental study conducted by Hintze et al. essentially administered both the DIBELS and CTOPP in March to 86 kindergarten students. The results revealed a moderate to strong correlation between DIBELS and CTOPP scores with the Initial Sound Fluency (ISF) subtest providing the highest predictive ability followed by the Phoneme Segmentation Fluency (PSF) subtest and no predictive ability demonstrated on the Letter Naming Fluency (LNF) subtest. These findings suggest that the DIBELS and CTOPP measure similar constructs when ORF is removed from the DIBELS test battery. It should be noted that in their study, Hintze et al. did not administer the ORF subtest of the DIBELS and administered all seven subtests of the CTOPP. In addition, Hintze et al. noted that although the criterion score established by DIBELS demonstrated a high level of sensitivity, it was at the expense of creating an inordinate number of false positives. That is, there were an inordinate number of kindergarten students with low DIBELS scores who did not perform poorly on the CTOPP.   When Hintze et al. adjusted the criteria scores for the DIBELS, the classification validity for poor readers increased significantly. These findings are consistent with conclusions drawn by Riedel (2007) indicating a historical consistency in DIBELS tendency to over identify false positives due to unreliable cut-off scores or criteria. Furthermore, Hintze et al. concluded that although DIBELS is a good measure of phonological awareness for kindergarten children, it significantly over identifies students as requiring intervention. Therefore, these researchers suggested that DIBELS should only be utilized as a screening measure and it is necessary to follow up DIBELS testing with superior predictive measures in order to appropriately prescribe RTI interventions and special education recommendations.

Finally, Tierney and Thome (2006) discuss the philosophical and practical implications of the DIBELS. More specifically, these researchers suggest that the national endorsement of DIBELS by the National Reading Panel has led to a phenomenon whereby that which DIBELS measures and teachers teach in American classrooms have become the same. Thus, Tierney and Thome suggest that DIBELS does not go beyond assessing what is directly taught and argue that it should also examine reading skills which are indirectly learned. Otherwise, these researchers warn that the DIBELS will serve to restrict and constrain how American teachers view and teach literacy. In fact, Tierney and Thome noted that a national “Corrective Reading” program has been launched which focuses exclusively on minute reading drills and literacy skills development and ignores other “rich” instructional skills such as oral story telling, writing, etc. Furthermore, these authors raise additional concerns that poor school districts receiving Reading First funds will focus exclusively on the five components of literacy while more affluent school districts will continue to engage students in richer, indirect reading instruction. Tierney and Thome also warn that DIBELS data are not being effectively utilized to guide reading instruction through data analysis, but simply identifying students at-risk for developing reading difficulties if specific and remedial reading intervention is not provided. In doing so, these authors believe that DIBELS perpetuates "scripted" teaching such as the Corrective Reading program and is swiftly eliminating creative teachers who engaged students in reading by incorporating rich and dynamic reading instruction into their lesson plans. Perhaps this is why some critics such as Pearson (2006, p. v) stated that "DIBELS is the worst thing to happen… since flashcards." Although Tierney and Thome predict that DIBELS may be effective in enhancing the short-term goals of reading instruction, they raise caution that it will not be effective in predicting and guiding intervention toward the enhancement of the long-term established goals of reading.

The collective body of DIBELS research presently reviewed raises specific questions which will hopefully be investigated further by the current research. More specifically, if the ORF (oral reading fluency) subtest from the DIBELS is indeed predictive of reading comprehension skills, other DIBELS subtests should be predictive of other specific processing skills involved in reading. Furthermore, if the DIBELS ORF subtest is so powerful at predicting reading comprehension skills (Riedel, 2007), then ORF alone or a combination of DIBELS subtest analysis should be able to powerfully discriminate between students that require and do not require intensive reading intervention programs, such as FFW-L. In addition, the present body of DIBELS research seems to indicate that DIBELS is simply measuring very basic and isolated reading skills which do not justly capture the Gestalt of reading ability. While this is a valid concern, the current research limited its focus to correlating DIBELS subtests to specific reading processes targeted by the FFW-L program (see Appendix C). Significant concern has been raised about DIBELS ineffectiveness at guiding instruction (Tierney & Thome, 2006) and serving simply as a mechanism for identifying at-risk students. The present research further investigated whether a DIBELS subtest analysis can meaningfully and significantly predict which students will benefit from an intensive reading intervention program. In fact, certain researchers (e.g. Hintze et al., 2003) have suggested that the cut-off scores for the DIBELS need to be readjusted in order to better predict true positives or students who are truly in need of intensive reading intervention. The present research also provided a direct opportunity to determine whether DIBELS cut-off scores are appropriate or require readjustment specific to the FFW-L intervention. Finally, certain researchers who investigated the DIBELS predictive ability related to reading skills purposefully excluded special education students from their sample (Riedel, 2007) with no explanation for their exclusion. Given that special education students typically require intensive intervention and the majority students with a learning disability have a specific reading disability, their inclusion in the present research is considered to be valuable and represents a more comprehensive sample.

Fast ForWord

Fast ForWord (FFW) is a computerized program initially developed and released in 1998 to improve speech and temporal processing skills in students specifically with language processing or speech impairments. However, FFW quickly gained support and popularity as an effective and practical reading intervention program for schools. In fact, Harrison and Gimbel (1998) designed a FFW training project prior to the release of FFW by the Scientific Learning Corporation with a local university and a hospital rehabilitation institute with the aim of making FFW more accessible to students in their school setting. The theoretical foundation for the efficacy of FFW impact on language processing and reading skills is consistent with Shaywitz's (1998) ground-breaking findings that severe reading disabilities, such as dyslexia, involve impairment in the occipito-temporal region of the brain. Although FFW has received significant and robust efficacy research support, cautious researchers (Rouse & Kruger, 2004) point out that the initial FFW efficacy studies were coordinated and supervised by SLC, the developer of FFW. Therefore, the validity of these initial efficacy studies can be questioned. In fact, the What Works Clearinghouse (WWC), which provides an online report of what interventions "work" based upon scientific data, conducted a review of the FFW program's efficacy data. An independent analysis of the WWC report was conducted by McArthur (2008). McArthur noted that 24 published and 91 unpublished studies were reviewed in the WWC report. The WWC ranks studies into the following three categories: 1) meets evidence standards, 2) meets evidence standards with reservations, and 3) does not meet evidence standards.

McArthur's analysis revealed that only six of the published and unpublished studies passed category 1 and category 2 established by the WWC. Furthermore, of these six studies, four were published by SLC and the other two studies did not pass peer review. McArthur also cautioned consumers to be aware that WWC quality control does not consist of independent peer reviews and allows the program developers to comment. Finally, McArthur critiqued the WWC report for its failure to include a study conducted by Cohen et al. (2005) in their review as McArthur considers this research study to be one of only a few that implemented the necessary experimental controls in their design.

SLC has conducted many of their own unpublished field studies as well as published studies addressing the benefits of the FFW-L program for developing reading skills. According to efficacy reports posted on SLC's website (n.d.) their initial clinical studies conducted from 1994 to 1995 involved controlled research designs and determined FFW to be efficacious in significantly improving the following skills: overall language skills, auditory processing speed, speech discrimination, phonemic and phonological awareness, grammatical and syntactic comprehension, overall language comprehension, and other receptive and expressive language skills. In 1996 a national field trial was conducted to determine whether the positive effects demonstrated in the clinical trials would generalize to real-world settings. This national field trial spanned 35 sites across the United States and Canada and involved the implementation of FFW in the following settings: a conventional clinic, private practice, school, or home setting. The target population consisted of students between the ages of 4 and 14 who exhibited difficulties with either listening or language comprehension skills. Results of this field trial, as reported by SLC, led the developers of SLC to conclude that students trained with FFW make, on average, a one to two year gain after four to eight weeks of training. However, SLC's report (n.d.) does not specify the skill areas where these gains were found. Instead, the report indicates that 90% of the students who participated in the national field trial made significant gains in one or more of the test areas. The outcome measures utilized in the field trial were described to "include" measures tapping the following skills: auditory word discrimination, following spoken directions, listening and speaking fundamentals, auditory processing speech, speech discrimination, language processing, grammatical comprehension, and overall language comprehension.

Prior to the public release of FFW, a school pilot study was conducted in 1997 involving nine school districts in four different states. The scope of this pilot study was to determine the efficacy of FFW training for students in grades kindergarten through 3rd grade who were determined by their teachers to be at-risk for failure in reading or language arts. The study involved more than 400 students who were randomly assigned to a treatment group (FFW) and a control group that was matched to the treatment group for age and gender. The 1997 school pilot study (n.d.) utilized the following two outcome measures for pre and post-test assessment: the Test of Auditory Comprehension of Language and the Phonological Awareness Test. The results indicated that prior to the FFW intervention, both groups were equally well below average on language comprehension measures with a mean percentile score of 12.5. Post testing revealed that the control group had improved these scores to the 21st percentile while the treatment group improved to the 49th percentile. Furthermore, 71% of the treatment subjects who received FFW training made an average improvement of 1.8 years and 75% of the treatment subjects were removed from the "at-risk" category. SLC also noted positive behavioral changes in the treatment group for the following: attention, cognitive flexibility, and distractibility. No mention was made regarding specific improvement in reading skills. Thus, it is likely that the significant benefits noted by SLC in their initial clinical, field, and pilot studies were exclusive to language-related skills.

More locally, an article posted in The Philadelphia Inquirer (2002, October 18) heralded a one-year improvement on state-based test scores of 180 points in math and 170 points in reading for the entire 5th grade student body in an elementary school where 90% of the students come from low-income families. The article also noted that the entire 5th grade student body for that particular school received FFW training during the one year interim. However, as this was not a published research study, other changes or interventions might have occurred in addition to FFW training and might have been partially or wholly responsible for the test score increases.

An independent review of SLC's studies was conducted by Wahl, Robinson, and Torgesen (2003). Although these reviewers recognized the general benefits of the FFW Language training program, Wahl et al. were uncertain as to benefits specific to reading outcomes such as phonemic decoding, word recognition, or reading comprehension. They also note that the research evidence provided by SLC does not specify whether the FFW-L training program has a unique instructional advantage when compared to conventional methods of direct instruction in reading and phonemic awareness skills.

Many research studies aimed at assessing the efficacy of FFW have been completed. However, according to McArthur (2008) only four studies to date provide randomized and controlled research data independent of SLC. These studies were conducted by Cohen et al. (2005); Hook, Macaruso, and Jones (2001); Pokorni, Worthington, and Jamison (2004); and Rouse and Krueger (2004). Additionally, an unpublished study by Borman and Rachub (2001 as cited in Rouse & Krueger; 2004) also provides an independent and randomized study of FFW efficacy.

Cohen et al. (2005) examined the effects of FFW in 77 children between the ages of 6 and 10 years, with severe mixed receptive-expressive language impairment. These researchers utilized a randomized controlled trial whereby all students who participated in the randomized control trials continued to receive their routine speech and language therapy during the FFW intervention. The subjects in this study were randomly assigned to three groups in which students: 1) received FFW intervention as home-based therapy for 6 weeks, 2) received computer-based games/ activities commonly used to promote language, or 3) did not receive any additional intervention beyond their typical speech and language therapy. Outcome measures involved assessment of language skills and did not include specific assessment of reading skills. The results indicated that although each of the three groups made significant gains in language scores, there was no additional effect or advantage for either of the two computer-based interventions over speech-language therapy.

In their study, Hook et al. (2001) examined the short and long-term (two years) effects of FFW-L on reading and spoken language skills for students identified with a reading disability using a value-added control-group design for students receiving a multisensory language and reading instruction during the two-year study period. One group (FWL) was also trained at the beginning of the study with the Fast FFW-L program, a second group (OG) initially received Orton Gillingham training, and a third group (LC) served as a longitudinal control group. All three groups received the regular multisensory language and reading instruction. The research findings revealed that both the FWL and OG groups made significant gains in phonemic awareness as compared to the LC group, but only the OG group made significant gains in word attack skills. None of the groups demonstrated any significant gains in word identification skills. Long-term effects measured by Hook et al. revealed that aside from the FWL group demonstrating significantly greater long-term gains in speaking and syntax measures, the three groups did not show any significantly greater gains in any of the literacy and reading measures at a two-year follow-up assessment. However, subjects in the FWL and OG groups did make equally significant long-term progress in phonemic awareness skills. These findings led Hook et al. to conclude that FFW-L program did not result in any additional or faster improvement on outcome measures. Additional analysis of the individual FFW-L tasks revealed that performance on the Circus Sequence and Phoneme Identification tasks of FFW were correlated with significant gains on the word identification subtest of the Woodcock Reading Mastery Test (WMRT), but not with the phoneme awareness subtests of the WMRT.

Pokorni et al. (2004) randomly assigned 60 students to one of three reading intervention groups. All subjects were nine years old, had speech IEPs, and were functioning at least one year below grade level in reading. Subjects were assigned to one of the following interventions: FFW, Earobics, and Lindamood Phonemic Sequencing (LiPS). The intervention took place during a 20-day summer program totaling 60 hours of intervention. The following pre-test and post-test measures were administered: two subtests from the Phonological Awareness Test (PAT) the CELF, and the Woodcock Language Proficiency Battery. The results revealed that the LiPS intervention was significantly better at improving subjects' ability to blend phonemes, but none of the three groups were found to be better at improving phoneme segmentation skills nor was significant improvement evidenced on any of the language or reading subtests. Furthermore, Earobics and LiPS were more effective than FFW in improving phonemic awareness skills. It should be noted that the LiPS intervention required a 1:4 teacher ratio due to the nature of the program; whereas FFW and Earobics are computer-based programs which do not involve interactive supervision and guidance from a teacher. Thus, one could argue that Earobics and FFW are much more prone to the negative effects of low motivation and indirect supervision than the LiPS program.

Rouse and Kruger (2004) conducted a randomized evaluation study which investigated claims that FFW significantly improves language and reading skills. More specifically, these researchers set out to determine whether computerized reading programs, such as FFW-L, actually improve student achievement. These researchers identified 485 students who were below the 20th percentile on state standardized reading tests and then randomly assigned them to the FWL treatment group or an untreated control group. Rouse and Kruger employed the following four outcome measures of language and reading skills in their study: 1) Reading Edge which was designed by SLC, 2) Clinical Evaluation of Language Fundamentals-3rd Edition (CELF-3), 3) a school-based assessment which incorporated the five essential components of reading identified by the NRP, and 4) percentile scores from the state standardized reading tests. Based on their results, Rouse and Kruger concluded that FFW-L program did improve some aspects of student's language skills, but these gains did not translate into broader improvement in reading ability among their large sample. Furthermore, these researchers noted that they essentially "stacked" the data analysis favorably for the FFW-L subjects by only including those students who met the "rules of completion" as established by FFW developers. That is, only those students who achieved a 90% completion on all FFW-L exercises or completed 80% on a majority of both sound and word exercises with notable steady progress were included in the final comparison group. More specifically, Rouse and Krueger noted that 45% of students actually met the rules of completion established by SLC. Despite having eliminated students from their study who did not successfully meet FFW criteria for successful completion of the program, these researchers still failed to find evidence for FFW-L efficacy in improving reading ability at a significantly higher level when compared to traditional school-based reading intervention programs. Similar findings were reported by Borman and Rachuba (2001), as cited in Rouse and Krueger (2004) which employed a similar design with 415 children in the Baltimore city public school system. Borman and Rachuba's unpublished findings failed to find any statistically significant difference between the FFW-L and control groups.

The collective body of FFW research presently reviewed raises specific questions which was investigated by the current research. More specifically, while prior researchers (Rouse & Krueger, 2004) purposefully readjusted the FFW rules of completion criteria when examining the effects of FFW on reading and language abilities and excluded students from the final sample who did not meet the readjusted FFW rules of completion, the present research examined and included students who participated in the FFW-L program regardless of their completion rate. In addition, the present research proposal included special education and non-exceptional students who participated in the FFW-L program regardless of their disability across kindergarten through third grade. Thus, the present research will provide a more comprehensive and realistic sample size. Finally, despite strong critiques of FFW's efficacy, the present research was based upon an underlying premise that the FFW-L program is indeed a research-based program which meets the standards of the NCLB criteria for scientifically-based intervention. Instead, the present research questioned whether DIBELS can appropriately discriminate between students who would and would not benefit from the FFW-L intervention. If indeed the temporal processing theoretical premise for FFW is valid and DIBELS is a good measure of specific reading processes that are also targeted by FFW (i.e., phonemic development and auditory processing), then data from the present study should reveal significant positive correlations between DIBELS and FFW-L variables.

Response to Common Literacy Intervention Programs

Despite ongoing debate as to whether the FFW training program provides any additional benefits to the development of reading skills in comparison to traditional reading intervention programs, the FFW programs are being widely used throughout the United States as a scientifically-based intervention to help struggling readers in elementary, middle, and high schools. Furthermore, the FFW training programs meet the criterion established by both the NCLB Act and the RTI model which stipulate that interventions need to be evidence-based by scientific research. In addition, research findings have provided general efficacy support for FFW as an intervention aimed at improving language and reading skills. Therefore, as with many intervention programs, it might prove more useful to investigate whether students are being appropriately referred to intensive reading intervention programs- specifically FFW in the present study. Such an investigation seems critical given the amount of time and money invested in scientifically-based programs such as the FFW, Lindamood, and Wilson Reading programs. Jenkins and Johnson (2009) noted that the success of the RTI model rests upon universal screening measures, such as the DIBELS, to assess which students are not benefiting from the instruction of the general education program (Tier 1) and are indeed in need of specific, evidence-based interventions (Tier 2) due to the assumption that without the provision of such intervention these students are at great risk of developing increasing academic difficulties. However, Jenkins (2003) warns that effective screening instruments for the RTI model of service delivery must satisfy three criteria. The first is classification accuracy, followed by efficiency, and third, consequential validity. Jenkins and Johnson also noted that screening instruments can be deemed to be accurate in two ways (true positives and true negatives) and can be incorrect in two ways (false positives and false negatives). Given the limited resources and space that most urban school districts have to invest in elaborate training programs such as FFW, it is critical that broad-based screening tools such as the DIBELS provide a significant degree of accuracy in properly identifying and recommending students for specific intervention programs. If not, students are in jeopardy of being immersed in an intensive intervention program that requires anywhere from 6-8 weeks of daily instructional time in order to be effectively implemented. The philosophy of the RTI model, as indicated by Jenkins and Johnson (2009), embraces a monitoring of progress during a minimal of a six week period based upon a prescription model of intervention services that rests primarily upon screening measures like the DIBELS. Recent research (Allor et al., 2001; Vaughn et al., 2009; Whiteley et al., 2007) has begun to investigate whether students do indeed "respond" or significantly benefit from prescribed Tier II interventions and has also developed terms such as "low responders" and "high responders" (Vaughn et al., 2009). But there are no published studies which specifically investigate the predictive validity of universal screening measures, such as the DIBELS, with successful completion of intensive reading intervention programs. That is, can the DIBELS subtests significantly predict which students will respond favorably to a costly Tier 2 intervention program such as FFW-L in order to minimize the number of false positives and false negatives as discussed by Jenkins and Johnson (2009)? By eliminating such non-responders and maximizing the amount of positive responders recommended to the FFW program for Tier II intervention, urban school districts such as Philadelphia can capitalize on their limited resources. Furthermore, this study also investigated whether a specific DIBELS profile exists unique to positive responders and non-responders. Identifying such profiles would greatly help to enhance school professionals' ability to interpret DIBELS scores in consideration for referral to reading intervention programs, specifically the FFW intervention program.

CHAPTER 3

METHODS

Participants

The investigator identified specific elementary schools in the Northeast Region that demonstrated satisfactory average participation levels on the FFW program to ensure that the data collected and analyzed reflected adequate student effort and participation on the FFW program. As learning to read involves different components and cognitive processes than later stages of reading, the study was limited to Kindergarten through 3rd graders. Limiting the sample to this grade range also helped ensure that students' records were available at the school at which they completed FFW, and that students were administered the DIBELS recently. Students enrolled at four elementary schools in the Northeast Region of the Philadelphia School District were selected for inclusion in the study if they met the following inclusion criterion: they had participated in FFW Language program while enrolled in K-3rd grade during the 2006-2007 and 2007-2008 school years. An original sample size of 203 subjects was obtained after matching DIBELS data to the original dataset of eligible students who participated in the FFW intervention. However, eight subjects did not have any DIBELS data and an additional 11 subjects only had DIBELS data for one of the six main DIBELS variables. Therefore, The final database included 184 students within four schools.

Procedures

Data Collection

The investigator had access to the FFW data set via the FFW Progress Tracker data set as a school psychologist employed at the Northeast Region of the Philadelphia School District. The investigator generated a list of students at each school that met the inclusion criterion and then generated a non-identifying participant numbers in order to code the data and remove any personally identifiable information (e.g., student names) under the Family Educational Rights and Privacy Act (FERPA)- see Appendix A.

The investigator then provided the Research Coordinator from the Philadelphia School District Office of Research with a list of students who met the inclusion criterion. The research coordinator then attached to this list data regarding students' DIBELS scores and demographic data as indicated in Appendix B. The research coordinator attached the non-identifying participant number for all data and removed original student identification numbers. Upon completion of the data collection and matching, the research coordinator provided the investigator with the final dataset.

Obtaining the Dataset.

No contact with students was required to conduct the study as the study was a retrospective analysis of data contained in students' school records. According to FERPA, parent permission for access to school records is only required if personally identifiable information (e.g., student names, student identification numbers, or student addresses) is released. Data collection for the study was conducted as described previously. The investigator did not administer any tests to students or have contact with the students included in the study. Students included in the study participated in the FFW Language intervention program as part of their school curriculum as determined by school teachers, staff, and administrators. All students included in the study participated in FFW prior to the present research investigation. The investigator was not part of the decision-making process for recommending the students at the elementary schools for participation in FFW. The investigator requested that the Research Review Committee of the Philadelphia School District grant the investigator permission to collect data from student records. Data collection was conducted by the investigator and the Research Coordinator form the Philadelphia School District Office of Research. The investigator and research coordinator collected information from student profiles on SchoolNet, EasyIEP, and FFW Progress Tracker (See Table 3.1 for FFW and DIBELS variable descriptions).

Table 3.1

List of FFW and DIBELS variable descriptors

FFW

DAYS	# of days participating in Fast ForWord program
PARTIC LVL	% score of participation level for Fast ForWord program
ATTEND LVL	% score for attendance level on Fast ForWord program
PERCENT COMP	% complete score for Fast ForWord program
FFSPERF	Successfully completing the program with 80% complete within 80 days or less
L CS	Circus Sequence Score (range 0-100) Language
L PI	Phoneme Identification Score (range 0-100) Language
L OMFF	Old McDonald Flying Farm Score (range 0-100) Language

Table 3.1 (continued)

L PW	Phonic Word Score (range 0-100) Language
L PM	Phonic Match (range 0-100) Language
L BC	Block Commander (range 0-100) Language
L LCB	Language Comprehension Builder (range 0-100)

DIBELS

ISF	Dibels Initial Sound Fluency raw score. The score is the number of initial sounds correctly identified in one minute.
ISF BM	Dibels Initial Sound Fluency Benchmark score. Student's level of fluency is either 1= established (at grade level), 2= emerging (needing strategic intervention), or 3= deficit (needing intensive intervention).
LNF	Dibels Letter Naming Fluency raw score. The score is the number of letters correctly named in one minute.
LNF BM	Dibels Letter Naming Fluency Benchmark score. Student's level of fluency is either 1= established (at grade level), 2= emerging (needing strategic intervention), or 3= deficit (needing intensive intervention).
PSF	Dibels Phoneme Segmentation Fluency raw score. The child receives one point for each phonemic sound of word produced in one minute (e.g. cat = k…a…t).
PSF BM	Dibels Phoneme Segmentation Fluency Benchmark score.: Student's level of fluency is either 1= established (at grade level), 2= emerging (needing strategic intervention), or 3= deficit (needing intensive intervention).
NWF	Dibels Nonsense Word Fluency raw score. The child is presented with a set of nonsense words. The score is the number of sounds/ nonsense words the child pronounces correctly in one minute.
NWF BM	Dibels Nonsense Word Fluency: Benchmark score. Student's level of fluency is either 1= established (at grade level), 2= emerging (needing strategic intervention), or 3= deficit (needing intensive intervention).
WUF	Dibels Word Use Fluency raw score. Assesses a child's ability to accurately use a provided word in the context of a sentence. The score is the number of words the child uses within one minute.
WUF BM	Dibels Word Use Fluency benchmark score. Student's level of fluency is either 1= established (at grade level), 2= emerging (needing strategic intervention), or 3= deficit (needing intensive intervention).
RTF	Dibels Retell Fluency score. Provides an assessment of ORF's validity. Helps to see whether child is comprehending and not simply reading fluently.
RTF BM	Dibels Retell Fluency Benchmark score. Student's level of fluency is either 1= established (at grade level), 2= emerging (needing strategic intervention), or 3= deficit (needing intensive intervention).
ORF	Dibels Oral Reading Fluency raw score. The child is asked to read a passage aloud. Pauses, omissions, and substituted words
ORF BM	Dibels Oral Reading Fluency Benchmark score. Student's level of fluency is either 1= established (at grade level), 2= emerging (needing strategic intervention), or 3= deficit (needing intensive intervention).
INST REC	Dibels Instructional Recommendation: based on student's results, this indicates if the student is 1= Benchmark (at grade level), 2= Strategic (needs some intervention), or 3= Intensive (needs intensive intervention).

Measures

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

According to the administration and scoring guide, the Dynamic Indicators of Basic Early Literacy Skills *(*DIBELS*)* assessments were designed to measure "critical skills that underlie early reading success" (Brunsman, 2005). Their purpose is to identify and monitor the progress of students who are unlikely to meet state reading standards in third grade. The administration and scoring guide asserts that DIBELS scores are good predictors of performance on high-stakes, summative tests of reading achievement. They can be used to evaluate the effectiveness of reading instruction both for individual students and for groups of students (e.g., school-wide). The DIBELS assessments were not meant to be used as the only diagnostic measures for students with reading disabilities. Although the DIBELS' administration and scoing guide manual (Good & Kaminski, 2002) does not specify the professional credentials or training requirements necessary for those administering the DIBELS, the requirements for training school personnel to administer and score the DIBELS are specific to individual states and school districts and can include paraprofessionals and regular classroom teachers. According to the Philadelphia School District's Fall, 2006 Final Budget Report, funds were allocated for literacy coaches to provide professional development to provide training to teachers to use handheld palm devices to administer the DIBELS (2006). Retrieved December 1, 2011 from http://www.phila.k12.pa.us/offices/cbo/fy07-final_budget_book.pdf. Furthermore, the Philadelphia School District also established a web-based training module for all DIBELS administrators (2006).

Retrieved December 1, 2011 from http://webgui.phila.k12.pa.us/offices/a/accountability/assessments/web-based-training-dibels--dra.

Students in kindergarten through Grade 3 complete three sets of benchmark tests each year. Although students who are identified as 'at risk' also complete progress-monitoring assessments throughout the school year, these assessments were not included in the current study. The following benchmark assessments were available: Letter Naming Fluency (kindergarten to Grade 1), Initial Sound Fluency (kindergarten), Phoneme Segmentation Fluency (kindergarten to Grade 1), Nonsense Word Fluency (kindergarten to Grade 2), Oral Reading Fluency and Oral Retelling Fluency (Grades 1 to 3), and Word Use Fluency (kindergarten to Grade 3). Letter Naming Fluency consists of a page of random upper- and lowercase letters. The student names as many letters as possible in one minute. For Initial Sound Fluency, the examiner reads four sets of four words (for a total of 16 items) and asks which word begins with a particular sound (e.g., /fl/). The student points to pictures representing the words to indicate the correct answers. Scores are reported as number of correct responses per minute. For Phoneme Segmentation Fluency, the examiner reads a word and then asks the student to say all of the individual sounds in the word. The maximum time allowed per sound is three seconds. The score is the number of correct sounds given in one minute. Nonsense Word Fluency consists of a page of two- and three-letter nonsense words (e.g., sig, rav, ov). The examiner scores the number of correctly pronounced sounds in one minute. For Oral Reading Fluency, the examiner listens while the student reads a fictional passage out loud and scores the number of correctly read words in one minute. Oral Retelling Fluency may be used with Oral Reading Fluency to check the relationship between oral fluency and comprehension. The examiner gives the student one minute to retell the details of the passage. The score for Oral Retelling Fluency is based on the number of words used by the student that indicate understanding of the passage. For Word Use Fluency the examiner says a word to the student and asks the student to 'use' the word. A response is correct if the word is used correctly in a phrase, expression, or sentence, or if the student gives the definition for the word (whether or not the word is included). The student's score is a count of words (all words, including the prompt word if used by the student) in correct responses in one minute.

The administration and scoring guide provides “instructional recommendations” for use with DIBELS scores. Decision rules are used to identify students who need intensive instructional support. The administration and scoring guide provides specific instructional suggestions for students who score below a percentile rank of 20 and between the 20th and 40th percentile ranks.

The administration and scoring guide contains a few indices related to reliability and validity, but the data provided are totally unusable because the developers do not describe the participants or methodologies used in the studies. The developers discuss indices related to alternate-form and internal consistency reliability, but do not mention interrater reliability. Three technical reports written by other authors are available on the DIBELS website and report concurrent correlations between the DIBELS and other measures. The developers have not described any studies investigating the predictive relationship of DIBELS scores to state assessments of reading standards.

Fast ForWord (FFW)

In order to assess student responsiveness to FFW, the following measures were selected for inclusion in the study: start date, last date, participation days, participation level, and the highest percent complete for each FFW Language activity. These measures can all be gathered from FFW Progress Tracker, which is an online data-analysis and reporting tool that enables educators to monitor and access individual progress on the FFW programs. The investigator had access to this information in his role as a school psychologist for the Northeast Region of the School District of Philadelphia. The start date and last date of participation provided information regarding the specific time period and total length of time each student participated in the FFW program. Participation days is a measure of the total number of days the participant actually worked on the FFW product. Participation days provided additional information regarding how often the child worked on the FFW program. Participation level is a score (percentage) indicating the degree to which the participant was meeting the selected protocol (Scientific Learning Corporation, 2006b). This measure provided information regarding how adequately the student adhered to the intervention protocol. Neuroscientific research demonstrates the importance of intensity of activity and frequency of activity in order to build and strengthen cognitive skills (Scientific Learning Corporation, 2006a). FFW studies have shown that maximum benefits are achieved if participants reach minimum participation goals of 80% (Scientific Learning Corporation, 2006a). Finally, the highest percent complete score (percentage) for each FFW activity indicates the highest amount of the content in that skill set that the participant mastered (Scientific Learning Corporation, 2006b). These measures provided information regarding which activities and skills proved more or less difficult for each student, and are measures of mastery over the skills targeted in each activity. Table 3.2 provides a comparison of skills assessed by the DIBELS and those skills targeted by the FFW intervention program. A detailed list of FFW Language activities and targeted skill areas can be found in Appendix C. Each activity targets various language, reading, and cognitive skills, such as listening accuracy, phonological awareness, auditory word recognition, listening comprehension, syntax and morphology, organization, word analysis, sound-letter correspondence, phonological fluency, English language conventions, focused attention, auditory processing, working memory, picture-word associations, sustained and focused attention, sequencing ability, and cross-modal processing. FFW studies have demonstrated that maximum benefit is achieved if participants reach minimum completion goals of 80% (Scientific Learning Corporation, 2006a). Research has demonstrated that participants with high participation levels and high completion levels on FFW products have demonstrated the most gains (Scientific Learning Corporation, 2006a). In addition to these measures, the investigator also recorded which product protocol each student completed. FFW Language has 50-minute, 75-minute, 100-minute protocols. The data collection form used to collect and record these FFW measures is contained in Appendix A.

Table 3.2

List of DIBELS and FFW targeted skill areas

	DIBELS	FFW
Skill areas assessed/targeted	Phoneme awareness Alphabetic principle Accuracy and fluency Vocabulary Comprehension	Phoneme discrimination Processing speed Sound sequencing Working memory Listening comprehension/sntax Auditory word recognition Sustained/focused attention

Research Questions and Data Analytic Procedures

The following research questions guided this inquiry.

1. Can the DIBELS significantly predict which students will reach the mastery level on the FFW Language program?

Heirarchical linear modeling (HLM) was employed to determine the predictive relationship between DIBELS scores and performance on the FFW Language program. HLM was selected over a hierarchical regression analysis due to HLM's sensitivity to statistical dynamics which are "nested" within existing group structures. For example, the present research study proposed to examine students' performance on outcome variables nested within the following groups: school, grade, and school year. Therefore, HLM would be able to take into account the fact that outcome scores may be affected by nested within group properties (i.e., school 1 student population is significantly different from school 2). In addition to detecting nested within group effects, HLM is also able to examine why the two groups differ (i.e., school 1 population has significantly more ESOL students than school 2).

Given the multilevel nature of the main research questions and the data available, hierarchical linear modeling or HLM was first attempted to analyze the data (Raundenbush & Byrk, 2002; Raundenbush et al., 2000). However, this approach was ruled out due to inconsistencies in the models obtained likely a result of limitations in the sample especially with regard to the number of groups available at the school level. A common guideline for estimating sample size for HLM is provided by Paterson and Goldstein (1991) who advocate a minimum of 25 groups at level 2 (number of schools) with at least 25 cases in each group at level 1 (number of students). However, considerable disagreement exists about criteria for sample size and many studies have been done with smaller samples. However, several have documented instability in models due to sample sizes lower than those recommended (Dickinson & Basu, 2005; Snijders, 2006; Okumura, 2007). As a result, multiple regression analysis was used to evaluate the data. The following DIBELS subtest raw scores represent the predictor variables for this analysis: ISF, LNF, PSF, NMF, WUF, RTF, and ORF (see Table 3.1). The Overall Percent Complete FFW variable served as the outcome or dependent variable. In addition, a secondary analysis was conducted using the “highest percentage completed” score for each of the seven individual FFW-L activities (see Appendix A) as the outcome variable instead of the Overall Percent Complete score. Finally, additional analyses were conducted in order to determine whether the DIBELS benchmark scores, rather than the raw scores, significantly predicted success on the FFW Language program using the overall FFW Percent Complete score as the outcome variable and then again using the Highest Percentage Complete score for each of the seven FFW-L activities as two separate hierarchical regressions.

2. Can the DIBELS significantly predict which students will meet the "rules of completion" for successful performance on the FFW-L?

For the same reasons noted above, HLM was not employed. Instead, a multiple regression analysis was conducted using the following DIBELS subtest raw scores: ISF, LNF, PSF, NMF, WUF, RTF, and ORF (see Table 3.1). The outcome variable was dichotomous representing those students who did or did not meet the established cut-off point for successful completion of the FFW-L program (overall Percent Complete score of 80% within 80 days or less). In addition, a secondary analysis which also employed a multiple regression was conducted in order to determine whether the DIBELS benchmark scores, rather than the raw scores, significantly predicted which students met or did not met the established cut-off point for successful completion of the FFW-L program. A final multiple regression analysis was conducted in order to determine whether the DIBELS raw score variable was able to predict successful completion for the seven individual FFW-L activities as per the established cut-off criterion.

3. Do demographic variables predict which students will perform successfully on the FFW Language program?

An ANOVA test was run to determine whether any demographic variables significantly predicted successful performance on the FFW-L program. The following demographic data (see Appendix B) represented the independent variables for this analysis: grade, school year, race/ethnicity, gender, age, and educational placement. The FFW-L Percent Complete score (see Table 3.1) served as the outcome or dependent variable.

CHAPTER 4
RESULTS

Analytic Approach

An initial database was generated by obtaining DIBELS data for those students in grades kindergarten through third grade in four elementary public schools located in the Northeast region of Philadelphia which demonstrated the highest overall levels of participation falling above the 70% cut-off criteria established by SLC and contained an adequate sample size of students who actually participated in the FFW intervention. Participation level refers to the extent to which a subject adhered to the chosen protocol (participation level). SLC recommends that students use the product for 50 consecutive minutes a day, five days a week. It essentially measures a student's time on-task on the days they show up at the FFW lab in school. For example, if a student logs off or stops working after 40 minutes on a 50-minute protocol or uses the program only four out of five days in a week, then his or her participation level would equal 80 percent. Due to signiifcant scheduling constraints in the Philadelphia public school's daily schedule which often render it difficult for students to be excluded from their regular instruction for 50 minutes every day of the week, the 70% cut-off criteria established by SLC seems to be practical to the current school population. Therefore, an original sample size of 203 subjects was obtained after matching DIBELS data to the original dataset of eligible students who participated in the FFW intervention. The original dataset was evaluated to decide which variables to maintain and to determine criteria for dealing with problem cases such as those variables with large amounts of missing data. Eight subjects did not have any DIBELS data and an additional 11 subjects only had DIBELS data for one of the six main DIBELS variables. Therefore, the final database included 184 students within four schools. Finally, three of the six main DIBELS variables contained in the original dataset were eliminated due to the complete absence of data (WUF) or too few data within the specific DIBELS variables (LNF and RTF). Examination of the dataset provided by the Philadelphia School District for the purposes of this study included data fields for the six main DIBELS variables for students in grades kindergarten through 3rd grade. The complete lack of data for the WUF variable suggests that this subtest was simply not administered or dropped from the school district's DIBELS testing procedure. It should be noted that according to the DIBELS 2008 manual, the WUF variable did not yet have established benchmark goals associated with the WUF raw scores. Furthermore, the dataset provided by the Philadlephia School District had very few entries for the RTF (n=11), and LNF (n=16) DIBELS variables. According to the DIBELS 2008 manual, the RTF variable did not yet have established benchmark goals associated with the RTF raw scores. Additionally, the LNF subtest is only administered to students in kindergarten and the first grade. Descriptive and correlational analyses are presented first followed by the results obtained from an ANOVA and multiple regression analysis.

Descriptive and Correlational Results

Percentages for the demographic variables can be found in Table 4.1. The descriptive results suggest that the present sample is representative of this urban public school setting within the United States, with the exception of an overrepresentation of Latino students and underrepresentation of African-American and Asian-American students. In addition, this sample of students receiving a targeted intervention had a higher proportion of identified special education students than the population as a whole.

Table 4.1

Descriptive Statistics for Demographic Variables

	N	%	NE Region %	(SDP %)
Gender				
Male	96	52.2	51.8	(51.2)
Female	88	47.8	48.2	(48.8)
Ethnicity				
European American	25	13.6	29.0	(13.2)
African American	74	40.2	31.1	(58.3)
Latino	84	45.7	24.0	(18.3)
Asian American	N/A*	N/A*	9.5	(7.1)
Other	1	0.5	6.4	(3.1)
Age				
5	2	1.1		
6	15	8.2		
7	46	25.0		
8	73	39.7		
9	46	25.0		
10	2	1.1		
Grade				
K	7	3.8		
1	14	7.6		
2	59	32.1		
3	104	56.5		

Table 4.1 (continued)

Educational Placement				
Spec Ed	54	29.3	13.4	(14.7)
Reg Ed	130	70.7	86.6	(85.3)
Primary Disability				
Non-Exceptional	129	70.1		(80.0)
SLD	31	16.8		(8.3)
Speech	16	8.7		(2.2)
Intellectually Disabled	3	1.6		(1.8)
Autism	1	0.5		(1.2)
Hearing Impaired	1	0.5		(0.1)
Emotional Disturbance	2	1.1		(1.6)
Gifted without Disability	1	0.5		(3.2)

** Data provided by SDP did not include a category for Asian ethnicity*

The sample consisted of 85.9% minority students with the greatest pool of subjects (56.5%) drawn from the third grade population. Males outnumbered female subjects by 3.4% and 29.3% of the sample received some level of special education services. Of those subjects receiving special education services, more than half (57.4%) had a primary diagnosis of a Specific Learning Disability and 29.6% had a primary diagnosis of a speech disability.

Means and standard deviations for the predictor and outcomes variables can be found in Table 4.2.

Table 4.2

Descriptive Statistics for Predictor Variables

	N	M	SD	Range
Predictor Variables				
DPSF	54	31.57	22.08	
DNWF	58	27.5	18.29	
DORF	166	37.86	32.51	
DINSTSREC	184	2.61	0.72	1 - 3
DPSFB	54	1.56	0.82	1 - 3
DNWFB	58	2.10	0.85	1 - 3
DORFB	166	2.71	0.60	1 - 3
Outcome Variables				
FFPERCOMP	184	52.17	22.05	0 - 100
FFPART	184	92.75	10.07	0 - 100
FFATTEND	184	58.93	11.87	0 - 100
FFDAYS	184	49.13	27.75	
FFSPERF	184	1.22	1.11	1 - 2
LCS	184	26.46	28.81	0 - 100
LPI	184	41.01	19.82	0 - 100
LOMFF	184	47.22	26.66	0 - 100
LPW	184	55.46	35.12	0 - 100
LPM	184	65.22	34.22	0 - 100
LBC	184	63.75	27.46	0 - 100
LLCB	184	69.01	34.41	0 - 100

Refer to Table 3.1 for a detailed description of above abbreviated variables.
Note. DPSF= Dibels Phonemic Segmentation Fluency raw score , DPSFB= Dibels Segmentation Fluency Benchmark score, DNWF=Dibels Nonword Fluency raw, DNWFB= Dibels Nonword Fluency Benchmark, DORF=Dibels Oral Reading Fluency raw, DORFB= Dibels Oral Reading Fluency Benchmark, DINSTREC= Dibels Instructional Recommendation, FFPERCOMP= Overall Percent Complete score on FFW program, FFPART=Percent Complete score for participation level on FFW program, FFATTEND= Overall Percen score for Attendance level, FFDAYS= Number

of days student spent on FFW program ,FFSPERF= Successful Performance level on FFW program, LCS=Circus Sequence Score, LPI= Phoneme Identification Score, LOMFF= Old McDonald Flying Farm score, LPW= Phonic Word Score, LPM=Phonic Match Score, LBC=Block Commander, LLCB=Language Comprehension Builder.

In addition, Pearson Product Moment Correlations were conducted to determine the relationships between predictor and outcome variables. Only two significant, positive correlations were found between the DIBELS Oral Reading Fluency (ORF) raw score and the Fast ForWord Percent Complete Variable (FFPERCOMP; $r=0.33$, $p<.01$) as well as DIBELS Oral Reading Fluency (ORF) raw score and the Fast ForWord Successful Performance score (FFSPERF; $r=0.18$, $p<.05$). Correlations between predictors and outcomes can be found in Tables 4.3 and 4.4.

Table 4.3

Pearson-Product Moment Correlations to determine the inter- relationship between DIBELS predictors and FastForWord outcome scores

Pooled Within-Groups Matrices

	DPSF	DPSFB	DNWF	DNWFB	DORF	DORFB	DINSTREC
FFPERCOMP	-.03	-.15	-.08	-.12	.33**	-.11	-.03
FFPART	-.13	.07	-.09	.01	.07	-.03	-.00
FFATTEND	-.01	.01	-.13	.19	.08	.05	.11
FFDAYS	.01	.10	-.01	.00	.01	.04	.01
FFSPERF	-.08	-.13	-.02	-.02	.18*	-.08	-.07
LCS	-.12	-.07	-.11	-.10	.17	-.03	-.02
LPI	-.21	-.03	-.14	-.12	.16	-.04	.02
LOMFF	-.04	.04	.07	.02	.14	-.07	.04
LPW	.17	-.27	.02	-.15	.28**	-.07	-.02
LPM	-.06	-.10	-.08	-.13	.30**	-.10	-.02
LBC	-.04	-.14	-.19	-.05	.30**	-.15	-.08
LLCB	.03	-.13	-.04	-.09	.30**	-.11	-.05

**p < .05. **p < .01. ***p < .001.*
Note. DPSFB, DNWFB, DORFB, and DINSTREC variables represent reverse scores
See Table 3.1 for variables' full name

Table 4.4

Pearson-Product Moment Correlations to determine the intra-relationship among DIBELS scores

Pooled Within-Groups Matrices

	DPSF	DNWF	DORF	DINSTREC	DPSFB	DNWFB	DORFB
DPSF	--------	.66*	.48**	-.25	-.57**	-.16	-.50**
DNWF		--------	.66**	-.33*	-.27*	-.39**	-.58**
DORF			--------	-.69**	-.28	-.59**	-.70**
DINSTREC				--------	.27	.65**	.99**
DPSFB					--------	.58**	.26
DNWFB						--------	.47**
DORFB							--------

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

Note. DPSFB, DNWFB, DORFB, and DINSTREC variables represent reverse scores

See Table 3.1 for variables' full name

Table 4.5

Pearson-Product Moment Correlations to determine the intra-relationship among Fast ForWord scores

Pooled Within-Groups Matrices

	FFPERCOMP	FFPART	FFATTEND	FFDAYS	FFSPERF	LCS	LPI	LOMFF	LPW	LPM	LBC	LLCB
FFPERCOMP	--------	.04	.20**	.29**	.50**	.67**	.64**	.73**	.76**	.77**	.79**	.82**
FFPART		--------	.33**	.08	.04	-.07	.03	-.00	.00	.26**	-.02	-.01
FFATTEND			--------	.33**	.07	.13	.20**	.09	.16*	.24**	.09	.14
FFDAYS				--------	.11	.17*	.04	.25**	.21**	.35**	.11	.30**
FFSPERF					--------	.42**	.39**	.37**	.37**	.30**	.32**	.30**
LCS						--------	.50**	.48**	.35**	.43**	.42**	.39**
LPI							--------	.49**	.35**	.41**	.44**	.38**
LOMFF								-------	.42**	.54**	.44**	.50**
LPW									--------	.47**	.54**	.65**
LPM										--------	.57**	.52**
LBC											--------	.70**
LLCB												--------

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$
See Table 3.1 for variables' full name

Correlations examining the relationship between demographic variables and both predictor and outcome variables were also conducted. Significant correlations to note between demographic variables and predictor variables (DIBELS) included a significant, positive correlation between a subjects' grade of enrollment and Oral Reading Fluency raw score (r=.40, p<.01), indicating that as the grade increases, so does the number of words read correctly per minute. A significant, positive correlation was also noted between subjects' age and their Phoneme Segmentation Fluency Benchmark score (r=-.32, p<.05), indicating that as age increases, so does phoneme segmentation skills. A significant, negative correlation between a subjects' grade of enrollment and Instructional Recommendation score from the DIBELS (r=.27, p<.01) was found. Likewise, the subjects' age was negatively correlated with the Instructional Recommendation score from the DIBELS (r=.27, p<.01), indicating that the younger the student and grade of enrollment, the more intervention recommended.

A correlational analysis between select demographic variables (grade and age) and overall outcome variables (Fast ForWord) revealed some significant correlation. Grade level was significantly and positively correlated with attendance level (r=.21, p<.01) and percent complete score (r=.35, p<.01), indicating that as grade increases, so does the attendance level and number of activities completed. Likewise, the student's age demonstrated a significant, positive correlation with attendance level (r=.15, p<.01) and percent complete score (r=.29, p<.01), indicating that as age increases so does the attendance level and number of activities completed.

In addition to the overall outcome variables, some individual outcome variables demonstrated a significant correlation with the grade and age of subjects. The subjects' grade of enrollment was significantly and positively correlated with all of the seven indivdual FFW activities. The subjects' age was also significantly and positively correlated with all but one of the seven individual FFW activities (the Circus Sequence Score). This indicates that as age and grade increases, so does their performance on all or most of the FFW individual activities.

Table 4.6

Pearson-Product Moment Correlations to determine the relationship between Demographic variables and the Predictor and Outcome variables

Pooled Within-Groups Matrices

	GRADE	ETHNICITY	GENDER	AGE	SPECIAL ED
FFPERCOMP	.35**	.04	-.13	.29**	-.15*
FFPART	-.10	-.03	-.04	-.06	.03
FFATTEND	.21**	-.06	-.10	.15*	-.15*
FFDAYS	.05	.07	-.02	.01	-.29**
FFSPERF	.09	.01	-.08	.11	-.08
LCS	.19*	.13	-.25**	.14	-.01
LPI	.26**	.08	-.15*	.28**	-.14
LOMFF	.31**	.10	-.16*	.25**	-.16*
LPW	.30**	-.12	.03	.23**	-.15*
LPM	.28**	.09	-.11	.23**	-.18*
LBC	.23**	.01	-.04	.21**	-.03
LLCB	.26**	-.02	-.07	.21**	-.07
DPSF	.15	.14	.09	.22	.14
DNWF	..09	.08	.03	.19	.15
DORF	.40**	.06	.05	.16*	.14

Table 4.6 (continued)

DINSTREC	.27**	.12	-.10	.27**	.00
DPSFB	-.12	.06	-.18	-.32*	-.16
DNWFB	.28*	.14	.02	.15	-.15
DORFB	-.03	.02	-.14	.09	-.02

$p < .05$. $**p < .01$. $***p < .001$. $****p < .0001$.

Note. Gender coded as 1 (male) and 2 (female); Special Ed coded as 1 (Yes) and 2(No)

See Table 3.1 for variables' full name

ANOVA Results

A one-way analysis of variance (ANOVA) was conducted to determine the effect of the remaining demographic variables (gender, ethnicity, and special education status) on dependent variables related to DIBELS and Fast ForWord outcomes. A one-way analysis of variance (ANOVA) was also conducted on the demographic variables age and grade after these variables were reconstituted in order to determine whether such an anlsyis would result in any more significant findings for those produced by the Pearson Product Moment Correlations. Age was dichotomized into two groups with the first group comprised of students between the ages of five and seven and the second group comprised of students aged eight to ten. Grade was reconstituted so that kindergarten and first grade students were combined into one group while retaining second and third grade students as seperate groups to comprise three total groups for grade. Post hoc analysis for the significant variables consisted of conducting pairwise comparisons to determine which affected the dependent variable most strongly. The following significant differences were found.

Significant differences were found for gender on the dependent measures related to LCS [F (1,183) = 11.74, $p < .001$], LPI [F (1,183) = 4.02, $p < .05$], and LOMFF [F (1,183) = 6.31, $p < .01$]. Males displayed higher scores on LCS (M=33.23, SD=31.26) than females (M=19.07, SD=23.93). Males displayed higher scores on LPI (M=43.79, SD=18.79) than females (M=37.98, SD=20.55). Males also displayed higher scores on LOMFF (M=51.88, SD=29.14) than females (M=42.14, SD=22.75).

Significant differences were found for ethnicity on the dependent measures related to LPW [F (2,180) = 7.00, $p < .001$], and LBC [F (2,180) = 3.36, $p < .05$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for African American subjects (M=66.47, SD= 34.40) was higher on LPW than Latino subjects (M=46.52, SD=33.63). White subjects (M=51.08, SD=33.69) did not significantly differ from African Americans or Latinos. Post hoc comparisons also indicated a significant difference between mean scores on the LBC variable for white subjects (M=54.72, SD=29.67) and African American subjects (M=69.45, SD=23.45).

Significant differences were found for grade on the dependent measures related to DORF [F (2 ,163) = 19.84, $p < .001$], DINSTREC [F (2,181) = 17.34, $p < .001$], DPSFB [F (2,48) = 4.28, $p < .05$], FFPERCOMP [F (2,181) = 12.85, $p < .001$], FFATTEND [F (2,181) = 4.64, $p < .05$], LCS [F (2,181) = 3.92, $p < .05$], LPI [F (2,181) = 8.89, $p < .001$], LOMFF [F (2,181) = 9.55, $p < .001$], LPW [F (2,181) = 8.95, $p < .001$], LPM [F (2,181) = 7.48, $p < .001$], LBC [F (2,181) = 4.55, $p < .01$], and LLCB [F (2,181) = 6.51, $p < .01$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for kindergarten and first grade students (M=1.81, SD=0.98) was significantly lower on the DINSTREC variable than second (M=2.78, SD=0.49) and third (M=2.66, SD=0.68) grade students, indicating that a significanlty lower number of kindergarten and first grade students were recommended for intensive reading remediation. Third grade students (M=49.03, SD=34.94) scored significantly better on the DORFB subtest than did second grade students (M=19.03, SD=16.00). Post hoc tests for DPSFB could not be conducted due to the small number of subjects within each of the variable levels. However, mean scores for the

DPSFB variable indicated the following: kindergarten and first grade students (M=1.89, SD=0.81), second grade students (M=1.45, SD= 0.62), and third grade students (M=3.00, SD=unavailable due to an N of 1). Post hoc comparisons using the Tukey HSD test also indicated group differences among the FFW variables. Third grade students (M=58.83, SD=20.08) scored significantly higher on FFPERCOMP than kindergarten and first grade students (M=39.43, SD=23.45) and second grade students (M=44.98, SD=20.94). Third grade students (M=60.73, SD=10.73) also achieved a higher attendance level (ATTEND LVL) than kindergarten and first grade students (M=52.48, SD=15.49). Third grade students also demonstrated consistency in scoring significantly higher on all of the FFW individual activities as compared to students in kindergarten, first, and second grades. Results of the Tukey HSD test for the FFW individual activities can be found in Table 4.12. There were no significant differences between whites and African Americans on any of the individual FFW activities.

Age was significant for DORF [F (1,164) = 14.55, $p < .001$], DINSTREC [F (1,182) = 5.71, $p < .05$], DPSF [F (1,52) = 5.65, $p < .05$], DPSFB [F (1,49) = 6.28, $p < .05$], FFPERCOMP [F (1,182) = 13.23, $p < .001$], LPI [F (1,182) = 8.71, $p < .01$], LOMFF [F (1,182) = 9.88, $p < .01$], LPW [F (1,182) = 11.10, $p < .001$], LPM [F (1,182) = 4.91, $p < .05$], LBC [F (1,182) = 6.21, $p < .05$], and LLCB [F (1,182) = 7.28, $p < .01$]. Older children scored higher on DORF (M=43.50, SD=34.58) than younger children (M=22.69, SD=19.48). Conversely, older children scored lower on the DINSTREC (M=2.69, SD=0.64) than younger children (M=2.43, SD=0.84), thus resulting in recommnedations for more intensive reading remediation. Younger children demonstrated weaker performance on the DPSF (M=28.51, SD=21.81) and DPSFB (M=1.76, SD=0.76) than older children did on the DPSF (M=46, SD=46.89, SD=17.31) and DPSFB (M=1.11, SD=0.33). Older children completed a higher percentage of the FFW program as indicated by FFPERCOMP variable (M=56.31, SD=21.76) as compared to younger children (M=44.24, 20.54). Older children also performed consistenly better on the following FFW individual activities LPI (M= 35.16, SD=20.07), LOMFF (M=38, SD=24.34), LPW (M=43.83, SD=32.70), LPM (M=57.56, SD=34.97), LBC (M=56.86, SD=26.03) and LLCB (M=59.68, SD=35.44) than did younger children LPI (M= 44.06, SD=19.06), LOMFF (M=, 51.57, SD=26.86), LPW (M=61.51, SD=34.93), LPM (M=69.21, SD=33.27), LBC (M=67.34, SD=27.60) and LLCB (M=73.86, SD=32.97)

Special education was significant for FFPERCOMP [F (1,182) = 4.20, $p < .05$], FFATTEND, [F (1,182) = 5.49, $p < .05$], FFDAYS, [F (1,182) = 13.63, $p < .001$], LOMFF, [F (1,182) = 4.43, $p < .05$] and LPM [F (1,182) = 6.18, $p < .05$]. Special education students completed a higher percentage of the FFW program as indicated by FFPERCOMP variable (M=57.30, SD=19.97) as compared to regular education students (M=50.05, SD=22.60). Special education students also had a better attendance rate as indicated by FFATTEND (M=62.07, SD=10.70) as compared to regular education students (M=57.62, SD=12.13). Special education students spent more days on the FFW program as measured by FFDAYS (M=60.46, SD=29.41) than regular education students (M=44.42, SD=25.71). Special education students performed consistenly better on the following FFW individual activities LOMFF

(M=53.57, SD=27.76) and LPM (M=74.81, SD=31.08) than did regular education students LOMFF (M=44.58, SD=25.84) and LPM (M=61.24, SD=34.78).

Table 4.7

ANOVA Results for Gender

	df	*F*	*p*	*η2*	*Magnitude*
DPSF	1	0.53	.472	0.01	small
DNWF	1	0.10	.755	0.00	small
DORF	1	1.40	.239	0.01	small
DINSTREC	1	1.54	.216	0.01	small
DPSFB	1	1.80	.186	0.04	small
DNWFB	1	0.02	.903	0.00	small
DORFB	1	3.57	.061	0.02	small
FFPERCOMP	1	3.54	.061	0.02	small
FFPART	1	0.31	.579	0.00	small
FFATTEND	1	3.10	.080	0.02	small
FFDAYS	1	0.56	.456	0.00	small
FFSPERF	1	1.17	.280	0.01	small
LCS	1	11.74	.001***	0.06	small
LPI	1	4.02	.046*	0.02	small
LOMFF	1	6.31	.013**	0.03	small
LPW	1	0.32	.572	0.00	small
LPM	1	2.21	.139	0.00	small
LBC	1	0.66	.415	0.00	small
LLCB	1	1.07	.302	0.01	small

**p<0.05, **p<0.01, ***p<0.001; See Table 3.1 for variables' full name*

Table 4.8

ANOVA Results for Ethnicity

	df	*F*	*p*	*η2*	*Magnitude*
DPSF	2	1.29	.284	0.05	small
DNWF	2	0.56	.574	0.02	small
DORF	2	1.85	.161	0.02	small
DINSTREC	2	1.10	.336	0.01	small
DPSFB	2	0.36	.698	0.02	small
DNWFB	2	0.31	.738	0.01	small
DORFB	2	0.49	.614	0.01	small
FFPERCOMP	2	2.46	.088	0.03	small
FFPART	2	0.10	.913	0.00	small
FFATTEND	2	0.70	.498	0.01	small
FFDAYS	2	0.57	.568	0.00	small
FFSPERF	2	0.03	.972	0.00	small
LCS	2	0.97	.381	0.01	small
LPI	2	1.30	.274	0.01	small
LOMFF	2	1.24	.293	0.01	small
LPW	2	7.00	.001***	0.07	medium
LPM	2	1.96	.142	0.02	small
LBC	2	3.36	.037*	0.04	small
LLCB	2	2.18	.116	0.02	small

**p<0.05, **p<0.01, ***p<0.001*

See Table 3.1 for variables' full name

Table 4.9

ANOVA Results for Grade

	df	*F*	*p*	*η2*	*Magnitude*
DPSF	2	2.37	.103	0.09	medium
DNWF	2	1.83	.170	0.06	medium
DORF	2	19.84	.000***	0.20	large
DINSTREC	2	17.34	.000***	0.16	large
DPSFB	2	4.28	.020*	0.15	large
DNWFB	2	2.67	.079	0.09	medium
DORFB	2	1.92	.150	0.02	small
FFPERCOMP	2	12.85	.000***	0.12	medium
FFPART	2	2.90	.058	0.03	small
FFATTEND	2	4.64	.011*	0.05	small
FFDAYS	2	0.06	.941	0.00	small
FFSPERF	2	0.79	.457	0.00	small
LCS	2	3.92	.022*	0.04	small
LPI	2	8.88	.000***	0.09	medium
LOMFF	2	9.55	.000***	0.10	medium
LPW	2	8.95	.000***	0.09	medium
LPM	2	7.48	.001***	0.08	medium
LBC	2	4.55	.012**	0.05	small
LLCB	2	6.51	.002**	0.07	medium

**p<0.05, **p<0.01, ***p<0.001*

See Table 3.1 for variables' full name

Table 4.10

ANOVA Results for Age

	df	*F*	*p*	*η2*	*Magnitude*
DPSF	1	5.65	.021*	0.10	medium
DNWF	1	3.26	.076	0.06	medium
DORF	1	14.55	.000***	0.08	medium
DINSTREC	1	5.71	.018*	0.03	small
DPSFB	1	6.28	.016*	0.11	medium
DNWFB	1	0.03	.863	0.00	small
DORFB	1	0.00	.997	0.00	small
FFPERCOMP	1	13.23	.000***	0.07	medium
FFPART	1	1.72	.192	0.01	small
FFATTEND	1	2.26	.134	0.01	small
FFDAYS	1	0.24	.628	0.00	small
FFSPERF	1	0.95	.331	0.00	small
LCS	1	3.71	.06	0.02	small
LPI	1	8.71	.004**	0.05	small
LOMFF	1	9.88	.002**	0.05	small
LPW	1	11.09	.001***	0.06	medium
LPM	1	4.91	.028*	0.02	small
LBC	1	6.21	.014*	0.03	small
LLCB	1	7.28	.008**	0.04	small

$*p<0.05$, $**p<0.01$, $***p<0.001$

See Table 3.1 for variables' full name

Table 4.11

ANOVA Results for Special Education Placement

	df	*F*	*p*	*η2*	*Magnitude*
DPSF	1	1.27	.265	0.02	small
DNWF	1	0.55	.462	0.01	small
DORF	1	1.70	.195	0.01	small
DINSTREC	1	0.02	.898	0.00	small
DPSFB	1	1.11	.298	0.02	small
DNWFB	1	1.32	.256	0.02	small
DORFB	1	0.01	.931	0.00	small
FFPERCOMP	1	4.20	.042*	0.02	small
FFPART	1	0.10	.755	0.00	small
FFATTEND	1	5.49	.020*	0.03	small
FFDAYS	1	13.63	.000***	0.07	medium
FFSPERF	1	1.26	.263	0.01	small
LCS	1	0.24	.625	0.00	small
LPI	1	2.36	.126	0.01	small
LOMFF	1	4.43	.037*	0.02	small
LPW	1	3.76	.054	0.02	small
LPM	1	6.18	.014*	0.03	small
LBC	1	0.52	.472	0.00	small
LLCB	1	1.54	.216	0.01	small

$*p<0.05$, $**p<0.01$, $***p<0.001$

See Table 3.1 for variables' full name

Table 4.12

Tukey HSD Results for Grade and FFW Individual Activities

	K/1st Grade		2nd Grade		3rd Grade	
	M	*SD*	*M*	*SD*	*M*	*SD*
LCS*	22.14	32.11	19.05	24.21	31.53	29.72
LPI**	33.90	19.81	34.41	20.69	46.19	17.83
LOMFF**	32.76	24.60	40.10	24.30	54.17	26.31
LPW**	36.05	29.80	46.86	32.50	64.25	35.03
LPM**	46.24	38.96	58.47	31.99	72.88	32.41
LBC**	53.95	30.20	58.17	27.22	68.89	26.09
LLCB**	55.00	36.90	60.47	34.93	76.67	31.80

** 3rd grade category significantly differed from 2nd grade.*

*** 3rd grade category significantly differed from both K/1st grade and 2nd grade categories.*

See Table 3.1 for variables' full name

Predictive Results

A standard multiple regression analysis was used to evaluate the relationships between variables. Using standard regression analysis, all predictor variables for DIBELS raw scores scores that had an adequate sample size were entered in the equation (DPSF, DNWF, and DORF). An additional analysis was conducted for the DIBELS benchmark scores (DPSFB, DNWFB, DORFB, and DINSTREC). Table 4.13 lists the sample size for each of the predictor variables. A standard regression procedure allows the researcher to evaluate each variable in terms of what it uniquely adds to the ability to predict the dependent variable compared to the predictive potential of other variables in the model (Tabachnick & Fidell, 2006). Information from the standard regression analysis did not yield significant results in the ability for either DIBELS raw or benchmark scores to predict scores on the Fast ForWord reading program.

Table 4.13

Sample size for Predictor Variables

	n	*M*	*SD*
DLNF	16	32.19	17.34
DLNFB	15	1.80	0.77
DPSF	54	31.57	22.08
DPSFB	51	1.65	0.82
DNWF	58	27.50	18.29
DNWFB	55	2.22	0.85
DWUF	1	42.00	N/A
DWUFB	0	N/A	N/A
DRTF	11	36.55	12.57
DRTFB	0	N/A	N/A
DORF	166	37.86	32.51
DORFB	166	2.71	0.60
DINSTREC	183	2.62	0.72

See Table 3.1 for variables' full name

CHAPTER 5
DISCUSSION

The present study was designed to evaluate whether the DIBELS, a universal screening measure commonly used to assess reading skills, predicts which students would benefit from FFW, an intensive reading intervention program implemented widely by a large, urban public school system. This was accomplished by employing a retrospective analysis of archival data from the 2006-2008 school years that were collected by the School District of Philadelphia's IRB personnel. The need for a prescriptive model for intervention is highly supported and emphasized by the "Response to Intervention" literature (Burns et al., 2006). While researchers suggest that universal procedures should be quick and easy to administer and interpret (Brunsman, 2005; Good & Kaminski, 2002; Jenkins & Johnson, 2009), research (Hintze et al., 2003; Nelson, 2008) has demonstrated that a screening battery which uses multiple measures provides better classification accuracy for distinguishing at-risk and typically developing students than any single-measure screening approach. Nonetheless, many school districts across the nation rely on single-measure assessments, such as the DIBELS, with an estimated 1,800,000 students being assessed yearly using the DIBELS (Samuels, 2007). Furthermore, those who support the use of DIBELS as a progress monitoring tool for the development of reading skills (Brunsman, 2005; Good & Kaminski, 2002; Nelson, 2008; Reidel, 2007) also suggest that DIBELS is not only useful at predicting future reading difficulties but also effective at identifying students in need of intervention. Interestingly, some of the proponents of DIBELS (Brunsman, 2005; Nelson, 2008;

Reidel, 2007) along with other researchers (Bishop, 2003; Hintze et al., 2003; Samuels, 2007; Tierney & Thome, 2006; Torgesen & Rashotte, 1999) respectfully acknowledge the disadvantages and potential risks of using the DIBELS as a sole measure of reading skill development including: the test's weak statistical power, invalid instructional categories leading to over-identification of students in need of reading intervention, poor concurrent and predictive validity with other well-known and established reading assessment measures, lack of sensitivity in measuring reading processes in early elementary grade students, and the lack of cohesiveness among the DIBELS individual subtests as demonstrated by ORF emerging as the only DIBELS subtest shown to possess some predictive power of future reading comprehension skills. Although the RTI methodology has been hailed for stimulating the provision of early reading intervention during Tier 1 and Tier 2 prior to a student being formally evaluated for special education services (Tier 3), there are currently no published studies available which specifically investigate the predictive validity of universal screening measures with the successful completion of intensive reading intervention programs. That is, once a student is identified by a universal screening measure, such as the DIBELS, to be in need of more intensive reading remediation available at the Tier 2 level, is the screening measure also able to accurately predict which of these students will benefit from a specific type of reading intervention. The ability to predict which students will respond favorably to a specific intervention serves as the primary scope of the current study as urban school districts such as Philadelphia are in dire need of capitalizing on their limited resources available for providing intensive reading interventions to early readers.

The present study is also unique in that the sample included special education students; whereas, prior research (Riedel, 2007) assessing the DIBELS ability to predict positive responders to specific reading intervention programs purposefully excluded special education populations from their sample. The current study was therefore guided by three major research questions. First, what are the relationship between demographic variables and students' performance on the FFW-L reading intervention? Second, can DIBELS directly and significantly predict which students will succeed and achieve mastery of reading skills as measured by the FFW-L indicators? And third, can DIBELS significantly predict which students are likely to excel in the FFW-L reading intervention program as defined by meeting the "rules of completion" established by SLC?

Descriptive Data Relative to Demographic Variables

Gender

The study's population of male and female students was consistent with population demographics for gender for both the Northeast Region and the overall School District data.

Ethnicity

Based upon School District data for both regional and overall district figures, African American students were underrepresented as compared to the overall School District population. Within the Philadelphia School District, African American students represent 58% of the student population; yet African American students represented 40% of the study population. This underrepresentation is consistent with the geographical population from which the study sample was gathered (Northeast Region of the SDP). African American students attending school in the Northeast Region make up 31% of the Region's student population. While the present study 's Caucasian population of 13% mirrored that of the Philadelphia School District's overall population of Caucasian students, Caucasians were grossly underrepresented based upon Regional data which reflect Caucasian students as making up 29% of the student population. This discrepancy within regional expectations is noteworthy when one considers that the present study did include notably more (by 9%), although not at a significant level, African American students than expected for the Northeast Region and almost twice as less Caucasian students than expected. This suggests that African American students were more likely to be referred for reading intervention than Caucasian students. Latinos were grossly overrepresented in the present study (46%) according to both School District and Regional population data. Latinos represent 18% of the School District's total population and 24% of the Northeast Region's student population. This extremely high level of overrepresentation could be explained by two possible factors. First, demographic data provided by the SDP for the study did not include any Asians in the final sample. It is unclear whether none of the 10% of Asian students that make up the Northeast Region participated in the FFW program during the study's target years or if the data provided by the SDP failed to designate students under the category of "Asian". Regardless, even if the latter explanation were valid and Asians were inadvertently categorized under other ethnicity this would still result in a significant overrepresentation of Latino students in the study's population according to both Regional (24%) and overall School District (18%) population data for Latinos. Therefore, the second explanation for the overrepresentation of Latino students is likely more valid. That is, due to NCLB legislation mandating and strongly demanding that all students, regardless of their disability, ethnic background, or primary language, achieve reading proficiency on standardized, high stakes state tests (i.e. PSSA), it is very likely that Latino students who typically make up a large percentage of the English as Second Language Learners (ESOL) program are disproportionately designated for intensive reading intervention programs such as FFW in order to "flood" them with reading intervention in hopes of improving their reading proficiency.

Grade and Age

The 2nd and 3rd grades represented the largest groups in the study (32% and 57%, respectively), reflecting a gross overrepresentation in the sample. The smallest group of students in the study were Kindergarten students (3.8%) followed by 1st grade students (8.1%). This overrepresentation of children in the higher grades and under-representation of Kindergarten and 1st grade students suggest that intensive reading interventions are mostly provided to students no earlier than the 2nd grade. This trend in the data might be related to the fact that the NCLB progress monitoring, statewide tests are administered to grades 3 through 8. Therefore, it is possible that due to the limited availability of the FFW program within the Philadelphia School District, teachers and administrators opt to provide the intensive reading intervention program to those students being monitored by the state and whose performance on these high-stake tests are directly tied to annual state funding. Interestingly if this hypothesis is correct, then the practice of providing intensive intervention to the older students is in direct opposition to research demonstrating that early reading intervention during preschool and primary grades has a greater impact on the student's reading skills and is more cost effective in the long-term. (Neuman & Dickinson, 2006; Shaywitz, 1998). An analysis of the average age of students in the study was consistent with the grade-based analysis. That is, the majority of the students in the study (90%) were 7 to 9 years of age.

Educational Placement & Primary Disability

Although the majority of the students in the study were regular education students (71%), the total percentage of special education students (29%) in the study was significantly higher than the special education population for the School District (15%) as well as the Northeast Region (13%). This indicates that the Northeast Region is also utilizing its Tier 2 intensive reading intervention program (FFW) for special education students following Tier 3. This over-representation of special education students in the study population is supported by the notion that special education students are more likely to struggle with reading and require intensive reading interventions than regular education students. However, this overrepresentation of special education students also indicates that special education students who are already receiving specialized instruction through their

IEP are also being referred to participate in a daily reading intervention program (FFW) located in a lab outside of their special and regular education classrooms. This overreliance on utilizing the FFW program for students already receiving special eductaion services is consistent with claims made by the FFW program developers (SLC report, n.d.) that students trained with FFW make average reading gains of one to two grade levels after four to eight weeks of training, suggesting that the FFW program is superior to most other traditonal school-based reading intervention programs. The majority of the special education students included in the study sample either had a primary diagnosis of a Specific Learning Disability (57%) or a Speech and Language Impairment (30%). These figures are consistent with the research literature conducted by SLC (2004) which indicates that the FFW reading intervention program is most efficacious for students with speech and language processing problems or students with reading disabilities. Given that national figures indicate that 80% of special education students have a Reading Disability (Shaywitz, 2005), the special education students referred to receive the FFW intervention coincide with the targeted demographic population for the FFW program.

Relationships within Predictors and Outcome Variables

Pearson product moment correlations were calculated to determine the relationships within predictor and outcome variables. The present analyses were generally consistent with research that has demonstrated the DIBELS universal screening measure subtests do not consistently or cohesively correlate well with one another. This supports prior criticisms in the research literature (Brunsman, 2005) that the DIBELS test lacks statistical power and that the DIBELS manual contains insufficient relaibility and validity measures. In fact, when examining the correlation between DIBELS benchmark scores along with the overall DIBELS instructional recommendation score, an extremely strong and atypical positive correlation was found between a student's Oral Reading Fluency benchmark score and their overall DIBELS Instructional Recommendation score. It would appear that the ORF benchmark score and the DIBELS instructional recommendation score are mutually inclusive and essentially the same variable. This has likely led prior researchers (Riedel, 2007) to suggest that the ORF benchmark score is the only useful DIBELS score that should be interpreted or considered when reviewing DIBELS results. However, this also questions whether the DIBELS benchmark scores are statistically sound with regard to construct validity. Examination of the DIBELS raw score subtest correlations revealed a similarly significant, although not as extreme, trend which indicates that the Oral Reading Fluency raw score is more strongly correlated with the overall DIBELS instructional recommendation score than the Phoneme Segmentation Fluency or Nonword Fluency subtests. This statistical tendency combined with the extremely brief one minute DORF testing sample indeed makes one wonder whether DIBELS is a sound assessment even when its scope is limited solely as a screening measure. Furthermore, the DIBELS statistical consistency between benchmark and raw scores is also problematic as demonstrated by the fact the DIBELS raw scores correlate consistently lower with the overall DIBELS Instructional Recommendation score than do the benchmark scores. Additionally, the DORF benchmark score correlated much more strongly with the DIBELS Instructional Recommendation score than it did with the DORF raw score. Also peculiar was the fact that the DORF benchmark score correlated significantly with the DPSF and DNWF raw scores, but not with the benchmark scores from these respective tests.

Although Fast ForWord is designed to be a reading intervention program and not an assessment or screening tool, Fast ForWord does have automated progress monitoring assessments which provide both percentile scores for each of the seven individual skill building activities as well as composite scores for overall progress which function like benchmark scores. Examination of the statistical correlations among the Fast ForWord individual activities performance indicators revealed a very consistent and significant correlational cohesion among individual activities. Furthermore, each of the seven individual FFW activities correlated similarly, without exception, with the overall FFW Percent Complete score. This suggests that the Fast ForWord program is a statistically sound intervention and could still be considered a viable intervention provided that students are appropriately referred to the FFW program. Analysis of the FFW measure that assesses a student's successful completion of the intervention revealed that only 9% of the sample successfully completed FFW. None of the demographic variables analyzed in this study significantly predicted successful completion of the FFW program. The fact that only 9% of the sample successfully completed the FFW program indeed questions the ability of the DIBELS to accurately predict which students would benefit from an intensive reading intervention such as Fast ForWord.

Predictors of Positive Response to FFW Intervention

The final and chief research question investigating the predictive ability of the DIBELS was examined using a multiple regression analysis model. None of the DIBELS variables significantly predicted a student's performance on the Fast ForWord program. Thus, it was not possible to investigate whether a specific DIBELS profile existed unique to positive responders and non-responders of the Fast ForWord program. Because this is the first research study that analyzed the use of DIBELS as a predictive screening tool for referring students specifically to the Fast ForWord intervention program, there is no specific research literature to refer to as a comparison base. Instead, prior research has demonstrated that DIBELS does have some concurrent validity with other reading assessment measures such as the CTOPP (Hinze et al., 2003). Nonetheless, these same researchers strongly caution against using the DIBELS as the only measure for referring students to timely and costly reading intervention programs. Currently there are no published research studies, with the exception of the current study, that have specifically investigated the predictive validity of universal screening measures, such as the DIBELS, with the successful completion of intensive reading intervention programs, such as Fast ForWord. Since the Fast ForWord intervention program is recognized as a scientifically-based intervention that meets the criterion established by both the NCLB Act and the RTI model, it is being widely utilized in school systems across the United States. However, as the present research study indicates, large urban school districts will continue to be faced with the dilemma of whether they are effectively referring students to intensive reading intervention programs during an era in which these programs are costly and limited in number due to funding deficits. The present research findings, despite inherent limitations, strongly indicate that the DIBELS alone is not effective in providing large urban school districts with the predictive ability to wisely allocate precious and limited intervention programs to those students who would benefit most from the intervention. Thus, the present research study confirmed the concerns of prior researchers (Jenkins & Johnson, 2009; Riedel, 2007) that suggested the DIBELS may inadvertently squander meager resources by over identifyinging students for intensive reading interventions.

Limitations

Limitations of the current study draw specific attention to the need for more rigorous and consistent methodological data collection whenever a large urban school district employs a standard universal screening measure, such as the DIBELS, to assess and target students in need of reading intervention strategies. The present database did not contain enough data for half of the DIBELS subtests due to missing data and/or a systematic altercation in the DIBELS assessment procedure by the school district examined in this study. Therefore, it is possible that the present study's inability to identify significant predictor variables was limited by this lack of data. However, it should be noted that two of the three DIBELS subtests (WUF and RTF) that had very little or completely missing data for the present dataset are noted, according to the DIBELS manual, did not yet have established benchmark criteria at the time of the data collection. Additionally, the final variable (LNF) that was not included in the data analysis is only administered to kindergarten students and during the first marking period of first grade. Given that the present sample only included seven kindergarten and 14 first grade students, the number of reported LNF scores (16) suggests that the LNF subtest of the DIBELS was properly administered. However, the present sample simply consisted of too few kindergarten and first grade students, which represents yet another limitation of this study. Future studies should attempt to obtain a more balanced sample across grade levels, particularly given the fact that prior research findings (Riedel, 2007) suggest that the DIBELS ORF subtest alone is enough to predict reading comprehension skills for first grade students prior to the developmental stage when reading comprehension requires more than single word reading fluency skills.

An inherent limitation of any retrospective analysis is the tendency to obtain a very brief sample of data in such a limited time span, which may not be indicative of prior or future events. While this poses as a legitimate limitation to the current study, the scope of the present study was inherently designed to measure the effectiveness of a very brief, universal screening measure to predict which students would benefit from and successfully complete a specific and intensive reading intervention program. Therefore, this limitation had little, if any, impact on the overall results of this study.

Although the original dataset was quite large with respect to the amount of students who had DIBELS test scores, the outcome variable resulted in a significant reduction of the final available dataset due to the limited number of "spots" available at each school for providing students with the Fast ForWord intervention program. Therefore, the manner in which specific students were referred to the Fast ForWord program may have been contaminated by other factors (i.e., teacher bias, parental demand, administrative and political influence, etc.) which may have resulted in a biased or nonrepresentative sample population. In fact, the dataset yielded in the present study did result in an underrepresentation of kindergarten, first grade, and African American students and a significant overrepresentation of Latino students. According to Buros (2010), no evidence could be located in the DIBELS manual that addressed the extent to which participants reflected U.S. demographic characteristics.   Therefore, it is impossible to compare the current study's demographic composition with that of the original DIBELS database. Nonetheless, although the present study did not mirror the local population statistics with regard to ethnicity, it may indeed serve as a real-life portrayal as to which specific ethnicities are being targeted for intervention based upon the aforementioned administrative and political factors. Interestingly, an analysis of the mean DIBELS raw scores across ethnicity revealed an unexpected trend whereby the African American students scored consistently higher on all of the DIBELS raw subtest scores than white and Latino students, with white students consistently obtaining the lowest score of all three ethnic groups. Closer examination of this trend revealed yet another limitation of the current study. That is, the number of subjects at each grade level was disproportionate with respect to ethnicity. More specifically, white students exposed to the FFW intervention program in kindergarten outnumbered African American students four to one, and Latino students two-to-one. White students also outnumbered African American students two-to-one in first grade, but Latinos outnumbered white students just slightly over a two-to-one ratio. In second grade, the ratio of white students exposed to the FFW intervention program declined dramatically with African American students outnumbering white students approximately three-to-one and Latino students outnumbering white students almost four-to-one. An even greater decline was noted in third grade with white students being outnumbered in their exposure to the FFW intervention program seven-to-one as compared to African American students and Latinos. Therefore, not only is the present study limited in its ability to make inferences specific to ethnicity effects, it also raises some overall concerns regarding potential cultural or ethnic issues which may be underlying certain aspects of how early certain students are exposed to an intensive reading intervention program such as Fast ForWord.

Implications for Future Research

The current study attempted to enhance the literature research base beyond the practice of simply utilizing a standard universal screening measure for identifying poor readers by enhancing the focus of such research onto how capable such screening measures are at predicting which students would benefit most from a specific intervention. This shift in focus is paramount given the recent shift to the RTI model in the educational realm. Thus, it not only seems logical but necessary that within such a response-oriented model that researchers and educational consumers alike demand that universal screening measures aimed at assessing students deemed at-risk for reading deficits also be designed to predict which students would benefit from specific types of intervention programs. Currenlty, the RTI literature base focuses exclusively on how well curriculum-based screening measures, such as the DIBELS, are able to accurately assess poor readers who are at-risk and in need of further intervention. Despite repeated studies which illustrate how such universal screening measures tend to significantly over identify at-risk readers, researchers and educational administrators continue to favor and advance the notion of utilizing a very meager screening system over that of a more concentrated and thorough evaluation of a student's reading skills through the use of multiple assessment measures. Therefore, resulting in a disproprtionate number of students being recommended for intensive reading intervention in the primary grades when students' motivational levels and stamina are vulnerable and educational resources are severely limited as is the case in typical urban school settings. In fact, the DIBELS test creators are currently releasing the next edition of DIBELS, DIBELS-Next, which aims to convert raw scores to percentile scores while maintaining the majority of the original subtests. Future research would better serve the educational community and its students by redirecting research objectives at comparing current universal screening measures with that of more thorough assessment practices aimed at measuring reading skills and related cognitive processing skills such as those employed by school psychologists. Although such an agenda shift in research clearly goes against the RTI tide currently surging through the field of educational research, courageous researchers within the field could easily challenge whether quick and efficient screening measures do indeed yield equal and/or better prescriptive data aimed at making recommendations for effective reading interventions specific to an individual student's needs. This could be achieved by simply expanding the current research design to include a secondary level of data that allows for the simultaneous analysis of both universal screening measures (i.e., DIBELS) with more comprehensive evaluation procedures such as the data derived from a psychoeducational evaluation.

While the design and results of the present study did not indicate the need for a standardized pre-post test assessment of the FFW intervention's effectiveness, future research could investigate this further to confirm the notion that quality and statistical integrity of an initial assessment and subsequent identification of at-risk readers has a directly positive effect in predicting which reading intervention strategies would prove most beneficial for specific student profiles. Without the provision of thorough and valid reading assessment data, even an effective reading intervention program might prove insignificant as evidenced by the present study.

Given that RTI and universal screening measures seem to be dominating the current practice of assessing and delivering reading intervention to students across the United States, it would also be useful to expand the current research design to explore the long-term effects that universal screening measures have on a student's progression of reading skills by conducting a five or ten year follow-up study. That is, specifically explore whether there are any significant, long-term differences in the reading skills between those students that successfully completed the FFW program and those who did not. It is assumed that despite having been recommended to an intensive reading intervention program that appeared to be largely ineffective for most of the study participants that the RTI process would nonetheless have provided these students with additional reading interventions that would have ammeliorated those reading deficits that were not successfully resolved by their exposure to the FFW program. If indeed the initial screening and identification of reading skills is as important as suggested by the current researcher, then it is expected that a follow-up study would result in unfavorable reading progress for those students who did not respond favorably to the FFW intervention.

The present study contained a few limitations. Future research could seek to overcome these limitations by obtaining a more balanced sample of subjects across ethnic groups and grade levels while also attempting to ensure that all of the DIBELS subtests are administered according to the procedures outlined in the DIBELS manual. For example, future research could further investigate the underlying factors or reasons why certain ethnic groups, such as African Americans, seem to receive intensive reading interventions later as compared to other ethnicities. Finally, given that this is the first study which examined DIBELS ability to predict responsiveness to a specific and intensive reading intervention program replication studies would be necessary to confirm the present findings. Furthermore, the DIBELS test creators are releasing a new edition of the universal screening measure and future studies could examine the predictive ability of the lastest version.

Implications for Future Practice

The current study sought to determine whether a specific universal screening measure widely used in America, despite its poor statistical properties, is effective at appropriately prescribing students to intensive reading intervention programs. While the current author recognizes that DIBELS was originally meant to be used as a screening measure, it is apparent by both the marketing techniques employed by DIBELS founders and the literature review that DIBELS is often the sole measure employed by both large and small school systems as part of the RTI (or equivalent) Tier process involved in referring students for additional, more intensive intervention. Given its economic feasability and the ever-increasing cuts in state and governemt funding to education, this practice of relying heavily and solely on a quick and brief universal screening measure is expected to continue. Therefore, it is imperative that researchers continue to not only challange the efficacy of such practices but also attempt to modify or create a better screening measure. For example, the CTOPP is a statistically more sound reading assessment than the DIBELS but requires significantly more time to administer (approximately 30-60 minutes) as well as a qualified professional to administer the test appropriately to ensure valid test data. Researchers could attempt to revise the DIBELS or develop an entirely different screening measure which finds an appropriate balance between administration time and valid, meaningful test data.

The present study has demonstrated that in the real-life conditions of a large, urban public school system, DIBELS seems to have very little, if any, specific predictive ability to designate students to appropriate intervention programs. The reasons for this are numerous and likely not exclusively due to the DIBELS poor statistical properties, but include other factors related to school and politically-based issues. Regardless of the reasons, it is quite evident that a weak identification system for detecting at-risk readers will lead to inappropriate referrals to specific intervention programs, such as the Fast ForWord program. Subsequently, in a system of limited resources inappropriate referrals result in mediocre intervention results as demonstrated in the present study and can cause a school system to abandon a worthwhile intervention. This, in fact, did occur during the process of the present study. A majority of the schools in the Northeast Region of the Philadelphia public school system abandoned the FFW program in 2010 when the site-license for full on-site technical suppoprt ended and was not renewed by the School District of Philadelphia. The FFW labs were transformed into other reading intervention programs (i.e., ACHIEVE 3000 and Study Island). Thus, it is evident how a statistically poor universal screening measure employed to provide critical identification data inevitably results in a neverending cycle of abandoning and replacing established, scientifically proven interventions for the next wave of available, marketed reading programs.

Essentially, the fault can not be put on the test, per se, but rather on the administrators and policymakers who place so much faith and investment in a test which was not originally designed to accurately identify the specific reading needs of individual students. Given that American schools do not seem prepared to abandon the use of DIBELS or other commonly used universal screening measures, then in order to better serve the needs of students consumers of such universal screening measures should demand a better level of accountability from the universal screening test developers. That is, there should be a grading criteria or rating scale imposed upon a reading assessment's statistical ability to predict appropriate and successful reading interventions. For example, based upon the current study and real-life conditions of large, urban school districts, DIBELS would receive a low rating to predict or discern which students would and would not benefit from the FFW intervention program. If indeed RTI follows a medical model of delivery, then it stands to reason that educational consumers, like medical consumers, should demand data that rates how likely or well a particular test will identify and pinpoint educational "illnesses" or deficits. Likewise, educational consumers should then also demand data that rates how well a student will respond to a specific treatment or reading intervention program along with the potential side effects based upon the original diagnostic profile.

Of course, another option for future practice is for educators and policymakers to go against the current tide and rethink the utility and limitations of the RTI process. More specifically, we should recognize the potential pitfalls of RTI as a solitary model for measuring a student's response to intervention when it has been demonstrated that the commonly used RTI progress monitoring measures for reading are severely limited in their ability to appropriately prescribe intensive reading interventions. Perhaps this underscores the notion that clinical judgment in assessing processing strengths and weaknesses related to reading skills can never be overlooked or replaced by RTI screening measures. While RTI may be good at identifying students who might be in need of some intervention prior to the painstaking process of a complete special education evaluation (Tier 3), RTI's typical reliance on quick, universal screening measures often by unqualified assessors results in a superficial understanding of which specific processes are specifically in need of intervention (i.e., simultaneous processing, attention, memory, kinesthetic, behavioral, environmental). The RTI screening process also tends to over identify students as being at-risk for reading difficulties resulting in a "one-szie fits all" intervention which theoretically runs the risk of overexposing students to intervention and causing an unintentional side-effect such as burning-out a student with interventions and causing a dislike for reading due to ineffective and unnecessary overexposure. In an era where education champions scientific and evidenced-based driven data, it is counterintuitive not to include more sophisticated screening and assessment measures that are now available from the field of neuropsychology. School psychologists are uniquely qualified to meet the demands of an economically stressed educational system while maintaining ethical and valid assessment practices that would ensure a significantly more reliable method for appropriately referring students to limited intervention resources that would prove not only cost effective, but more importantly benefit individual students by providing them with appropriately matched and effective interventions during their initial remediation experiences so as to foster positive emotional attributes toward learning challenges and remedial instruction. As indicated by the literature review, research has demonstrated that a screening battery which utilizes multiple measures of reading skills provides better classification accuracy for distinguishing at-risk and typically developing students than any single-measure screening approach. Therefore, while current procedures within public school systems, as governed by IDEIA, tend to remove school psychologists from the critical role of the direct assessment of pre-intervention reading skills during Tier 1 and Tier 2 processes, perhaps state and government lawmakers should cease their practice of trivializing the importance of Tier 1 and Tier 2 screening measures and recognize that students deserve to receive a valid assessment that properly drives a prescription for an effective intervention plan similar to the diagnosistic and intervention practices driven by the medical model. At the very least, if school systems can not systematically and economically afford students the luxury of receiving a diagnostic and prescriptive reading evaluation during the Tier 1 and Tier 2 processes, then policy and law should require that the RTI team consult with the school psychologist to determine the diagnostic and prescriptive validity of individual test results before exposing a student to an intensive, costly, and time-consuming intervention program.

The current study is the first to investigate the technical and practical implications of using the DIBELS to predict which students will respond favorably to a specific reading intervention, namely the FFW intervention program. It is hoped that the information integrated in this report will stimulate some positive discussion and changes in the screening procedures currently being widely employed across American schools in order to better serve and educate American children that demonstrate symptoms of early reading deficits.

REFERENCES

AIMSweb RTI Assessment and Data Management (n.d.). Retrieved March 3, 2010 from http://www.aimsweb.com/

Allor, J.M., Fuchs, D., & Mathes, P. (2001). Do students with and without lexical retrieval weaknesses respond differently to instruction? *Journal of Learning Disabilities, 34*(3), 264-275.

Bishop, A. G. (2003). Prediction of first-grade reading achievement: A comparison of fall and winter kindergarten screenings. *Learning Disability Quarterly26* (3), 1-13.

Brunsman, B. (2005). Review of the DIBELS: Dynamic Indicators of Basic Early Literacy Skills 6th Edition. In Spies, R. & Plank, B. (Eds.), *The Sixteenth Mental Measurement Yearbook.* Lincoln, Nebraska: The Buros Institute of Mental Measurement.

Burns, M. K., Tilly, W. D., Gresham, F. M., Telzrwo, C. F., Shapiro, E. S., & Kovaleski, J. F. (2006). Response-to-intervention debate: Using data over dogma. *Communiqué, 34*(7), 5-6.

Buros Center for Testing. (2010). *Evaluating reading tests for the state of Florida: Dynamic Indicators of Basic Early Literacy Skills*, 6th edition (DIBELS) and the Florida Assessments for Instruction in Reading (FAIR). Lincoln, NE: Author. Retrieved from http://www.fcrr.org/FAIR/Buros Final%20Report-FLAReadingTestEvalsMay04-10.pdf

Cohen , J. (1998). Statistical *power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, W., Hodson, A., O'Hare, A., Boyle, J., Durrani, T., McCartney, E. et al. (2005). Effects of computer-based intervention through acoustically modified speech (Fast ForWord) in severe mixed receptive-expressive language impairment: Outcomes from a randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 48*, 715-729.

Deppeler, J. M., Taranto, A. M., & Bench, J. (2004). Language and auditory processing changes following Fast ForWord. *The Australian and New Zealand Journal of Audiology, 26*, 94-109.

Dickinson, L. Miriam and Anirban Basu. (2005). Multilevel Modeling and Practice-Based Research. *Annals of Family Medicine. 3*(1), 52-60.

Downing, J., & Kan Leong, C. (Eds.). (1982). *Psychology of reading*. New York, NY: Macmillan Publishing, Co., Inc.

Federal Register, 2006. Individuals with Disabilities Act, 2004.

Fiorello, C. A., Hale, J. B., & Snyder, L. E. (2006). Cognitive hypothesis testing and response to intervention for children with reading problems. *Psychology in the Schools, 43,* 835-853.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available: http://dibels.uoregon.edu/.

Harrison, S., & Gimbel, J.F. (1998). A Collaborative Computerized Language Training Project between Gonzaga University and St. Luke's Rehabilitation Institute. In: Coming Together: Preparing for Rural Special Education in the 21st Century. Conference Proceedings of the American Council on Rural Special Education (18th, Charleston, SC, March 25-28, 1998); see RC 021 434.

Hook, P. E., Macaruso, P., & Jones, S. (2001). Efficacy of Fast ForWord training on facilitating acquisition of reading skills by children with reading difficulties – a longitudinal study. *Annals of Dyslexia, 51*, 75-96.

Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *SchoolPsychology Review, 32*(4), 541-556.

Jenkins, J. R. (2003, December). *Candidate measures for screening at-risk students.* Paper presented at the National Research Center on Learning Disabilities Responsiveness -to-Intervention symposium, Kansas City, MO. Retrieved March 2, 2010 from http://www.nrcld.org/symposium2003/jenkins/index.html

Jenkins, E., & Johnson, J.R. (2009, April 30). *Universal screening for reading problems: why and how should we do this?* Video webcast presented by the RTI National Online Forum. Article retrieved February 12, 2010 from http://www.rtinetwork.org/Essential/Assessment/Universal/ar/ReadingProblems

Kovaleski, J. F. (2006). IDEA reauthorization includes RTI: Now what? *Communique, 34*(3), 26.

Loeb, D. F., Stoke, C., & Fey, M. E. (2001). Language changes associated with Fast ForWord-Language: Evidence from case studies. *American Journal of Speech-Language Pathology, 10*, 216-230.

McArthur, G. (2008). Does What Works Clearinghouse Work? A Brief Review of Fast ForWord. Australasian Journal of Special Education, *32*(1),101-107.

Nelson, J. M. (2008). Beyond correlational analysis of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A classification validity study. School Psychology Quarterly, 23(4), 542-552.

Neuman, S., & Dickinson, D. (Eds.). (2006) *Handbook of Early Literacy Research* (Vol. 2). New York: The Guilford Press.

Okumura, T. (2007). Sample size determination for hierarchical linear models considering uncertainty in parameter estimates. *Behaviormetrika,34*(2), 79-94.

Paterson, L. and H. Goldstein (1991). New statistical methods for analysing social structures: an introduction to multilevel models. *British Educational Research*

Journal 17(4), 387-393.

Pearson, P. (2006). Foreword. In K. Goodman (Ed.), *The truth About DIBELS: What it is, what It does* (pp. v-xix).Portsmouth, NH: Heinemann. Retrieved on March 2, 2010 from http://www.u.arizona.edu/~kgoodman/books.html

Philadelphia School District. (2006). Final Budget Book. Retrieved December 1. 2011 from http://www.phila.k12.pa.us/offices/cbo/fy07-final_budget_book.pdf.

Philadelphia School District. (2006). Web-Based Tarining. Retrieved December 1. 2011 from http://webgui.phila.k12.pa.us/offices/a/accountability/assessments/web-based -training- dibels--dra.

Pokorni, J.L., Worthington, C.K. & Jamison, P. J. (2004). Phonological Awareness Intervention: Comparison of Fast ForWord, Earobics, and LiPS. *Journal of Educational Research, 97*(3), 147-157.

Ramus, F. (2001). Dyslexia – Talk to two theories. *Nature, 412*, 393-395.

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

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., & Congdon, R.T.,Jr (2000). *HLM 5: Hierarchical linear and nonlinear modeling.* (Statistical software manual). Lincolnwood, IL: Scientific Software International.

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary In urban first-grade students. *Reading Research Quarterly*, 42(4), 546-567.

Robeck, M. C., & Wallace, R. R. (Eds.). (1990). *The psychology of reading: An interdisciplinary approach* (2nd ed.) Hillside, NJ: Lawrence Erlbaum Associates.

Rouse, A. B.; Krueger (August 2004). Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program". *Economics of Education Review* 23 (4), 323–338.

Samuels, S.J. (2007). The DIBELS tests: Is speed of barking at print what we mean by reading fluency? Reading Research Quarterly, 42(4), 563-566.

Scientific Learning Corporation (n.d.). Retrieved March 3, 2010 from http://www.scilearn.com/results/

Scientific Learning Corporation. (2004). Improved language and reading skills by students in the school district of Philadelphia who were receiving services for special education and who used Fast ForWord products. *MAPS for Learning: Educator Reports, 8*, 1-4.

Scientific Learning Corporation. (2006a). *Fast ForWord® teacher's handbook: A resource guide*. Oakland, CA: Scientific Learning Corporation.

Scientific Learning Corporation. (2006b). *The power of Progress Tracker: Workbook.* Oakland,CA: Scientific Learning Corporation.

Shaywitz, S. (2005). *Overcoming Dyslexia: A New and Complete Science-Based Program for Reading Problems at Any Level***.** New York: Random House.

Shaywitz, S. (1998). Current concepts: Dyslexia. *The New England Journal of Medicine, 338*(5), 307–312.

Snijders, T.A.B., 2006. Statistical Methods for Network Dynamics. In: S.R. Luchini et al. (Eds.), *Proceedings of the XLIII Scientific Meeting, Italian Statistical Society*, pp. 281-296. Padova: CLEUP.

Snyder, S. (2002, October 18). District schools' test scores on a rise. *The Philadelphia Inquirer*.

Tallal, P. (2000). Experimental studies of language learning impairments: From research to remediation. In D. V. M. Bishop & L. B. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention, and outcome* (pp. 131-155), New York, NY: Psychology Press.

Tabachnick, B.G., & Fidell, L.S. (2006). *Using Multivariate Statistics* (5th Ed.). New York: Allyn &Bacon

Tierney, R., & Thome, C. (2006). Is DIBELS leading us down the Wrong path? In K. Goodman (Ed.), *The truth about DIBELS: what is it, what it does* (pp. 50-59). Portsmouth, NH: Heinemann.

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

Troia, G. A. (2004). Migrant students with limited English proficiency. *Remedial and Special Education, 25*, 353-366.

Troia, G. A., & Whitney. (2003). A close look at the efficacy of Fast ForWord Language for children with academic weaknesses. *Contemporary Educational Psychology, 28*, 465-495.

U.S. Department of Education, Office of Elementary and Secondary Education, 2001. *No Child Left Behind: A Desktop Reference,* Washington, D.C., 20202. Retrieved March 3, 2010 from http://www2.ed.gov/admins/lead/account/nclbreference/page.html

VanDerHeyden, A. M., Witt, J. C., & Nacquin, G. (2003). The development and validation of a process for screening and referrals to special education. *School Psychology Review, 32*, 204-227.

Vanderwood, M. L., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (2002). The contribution of general and specific cognitive abilities to reading achievement. *Learning and Individual Differences, 13*, 159-188.

Vaughn, S., Wanzek, J., Murray, C., Scammacca, N., Linan-Thompson, S., & Woddruff, A. (2009). Response to early reading intervention: Examininghigher and lower responders. *Exceptional Children, 75*(2), 165-183.

Wahl M, Robinson C, Torgesen J. (2003). Florida Center for Reading Research: Fast ForWord Language. Florida Center for Reading Research. Retrieved March 2, 2010 from http://www.fcrr.org/FCRRReports/PDF/FastForWordLanguage.pdf

Whiteley, M., Smith, C.D., & Connors, L. (2007). Young children at-risk of literacy difficulties: factors predicting recovery from risk following phonologically-based intervention. *Journal of Research in Reading, 30*(3), 249-269.

Wilkinson GS, Robertson GJ. *WRAT4: Wide Range Achievement Test professional manual.* Lutz, FL: Psychological Assessment Resources; 2006.

Appendix A
Data Collection Form to be completed by Mr. Cavallo (school psychologist in the Northeast Region of the Philadelphia School District)

Student ID #: ________

School: ____________________________

Year: _________________________

FAST FORWORD DATA		Source of data
FFW begin date		FFW Progress Tracker
FFW end date		FFW Progress Tracker
Participation Days		FFW Progress Tracker
Participation Level %		FFW Progress Tracker
Attendance Level %		FFW Progress Tracker
Protocol Used (50min, etc.)		FFW Progress Tracker

FFW Language Activity Completion

FFW Activity	CS Circus Sequence	PI Phoneme Identificat.	OMFF Old MacDonald's Flying Farm	PW Phonic Word	PM Phonic Match	BC Block Commander	LCB Language Comp. Builder
Highest % Complete							

Appendix B

Data Collection Form to be completed by the Research Coordinator form the Philadelphia School District Office of Research

Non-Identifying subject ID#: ____________

Variable	**Data**	**Source**
DEMOGRAPHIC DATA		
Grade during FFW participation		School Net
School Year		School Net
Race/Ethnicity		SchoolNet
Gender		SchoolNet
Age		SchoolNet
Disability Classification		Easy IEP
Educational Placement		Easy IEP

DIBELS DATA

Date Administered		Source: School Net

Bench-mark Tests	ISF	ISF BM	LNF	LNF BM	PSF	PSF BM	NWF	NWF BM	ORF	ORF BM	RTF	RTF BM	WUF	WUF BM	INS REC
Score															

Appendix C
List of FFW Language activities and targeted skill areas

FFW Task	*Activity description*	*Targeted language/reading skill*
Old MacDonald's Flying Farm	Clicks and holds the flying animal to hear a repeated sound. Releases the animal when the sound changes.	Phoneme discrimination; sustained and focused attention; processing speed
Block Commander	Follows verbal instructions to identify and manipulate objects of various colors and shapes.	Listening comprehension and syntax; working memory; processing speed
Circus Sequence	Identifies a sequence of sounds by clicking buttons that correspond to the sound sweeps.	Working memory; sound sequencing ability; processing speed
Phonic Match	Matches sounds represented by a grid of tiles by clicking on one tile and finding another tile with the identical word.	Auditory word recognition; phoneme discrimination; working memory; processing speed
Phonic Word	Clicks the picture that represents the object that the exercise instructs the player to identify.	Auditory word recognition; phoneme discrimination; working memory; processing speed
Phoneme Identification	Identifies the character that matches the target sound.	Working memory; phoneme discrimination; processing speed
Language Comprehension Builder	Clicks the picture that has the most accurate representation of the sentence.	Listening comprehension and syntax; working memory; processing speed