The Impact of Participation in Four Year Old Kindergarten on Future Literacy Skills

By

Nicholle Breikjern

A Thesis Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Education Specialist

School Psychology

At

The University of Wisconsin – Eau Claire

(November 2013)

Graduate Studies

The members of the Committee approve the thesis of Nicholle Breikjern presented on November 18th, 2013.

Dr. Mary Beth Tusing, Chair

Dr. Melissa Coolong-Chaffin

Dr. Cathy Thorsen

APPROVED:

University Dean of Graduate Studies

The Impact of Participation in Four Year Old Kindergarten in Future Literacy Skills

By

Nicholle Breikjern

The University of Wisconsin-Eau Claire, 2013
Under the Supervision of Dr. Mary Beth Tusing

The current study examined the short and long-term associated impact of participation in a school district supported four-year-old kindergarten (4K) program on early literacy skills and reading acquisition. Over 1100 children across five cohort years were included in the study. The early literacy skills of 4K participants and non-participants were assessed during their kindergarten year using the Dynamic Indicators of Basic Early Literacy Skills assessment system. In addition, student performance on the state mandated third grade reading assessment was also analyzed. Children who participated in the 4K program entered kindergarten with stronger fluency of letter names and phonological awareness, as compared to nonparticipants. Comparisons to a comparison group indicated that, during mid-year assessments, 4K participants demonstrated stronger letter name fluency, phonological awareness, and phoneme segmentation fluency. 4K participation was also associated with lower reading risk status, as determined by DIBELS Instructional Recommendations scores. Despite differences in early literacy performance upon kindergarten entry, 4K participants did not differ relative to nonparticipants on the reading portion of the third grade Wisconsin Knowledge and Concepts Exam. Findings are discussed in terms of the relationship between 4K and kindergarten literacy skills, the relationship between 4K and future

reading risk status, and the relationship between 4K and third grade reading assessments. Limitations and implications for future research are also discussed.

Mary Beth Tusing 12/2/13

Dr. Mary Beth Tusing, Thesis Advisor Date

TABLE OF CONTENTS

CHAPTER I

Introduction

Concerns about widespread deficits in children's literacy skills have been present for decades, and a great deal of research on the development and remediation of reading and writing skills has been completed by both government-based organizations and researchers alike (Adams, 1990; Chall, 1967; NICHD, 2000; Snow, Burns, & Griffin 1998). Findings paint a grim outlook for many children in America. In 1997, The National Institute of Child Health and Human Development reported, "40% of the U.S. population has reading problems severe enough to hinder their enjoyment of reading" (Grossen, 1997). Additional longitudinal studies show findings that more than one in six children experience reading difficulties in grades one through six (Kame'enui, 1996). According to the National Association of Educational Progress (NAEP) Nation's Report Card (2011), the number of children scoring below the "basic" proficiency in reading on fourth grade measures is lower than in the past decade; however, fourth grade performance has largely become stagnant and unchanging across measures for 2007, 2009, and 2011. Therefore, while these concerns began decades ago, they continue to ring true in the current educational culture.

While concerns for academic deficits are noted for children of all backgrounds, the so-called "achievement gap" is of specific concern. The achievement gap is defined as the discrepancy between children of ethnic or low socioeconomic backgrounds and children of median to high economic backgrounds on measures of reading achievement (NAEP, 2011). The NAEP report of 2005 suggested that from 1995 to 2005,

approximately 60% of students growing up in poverty failed to meet basic literacy levels mandated by the federal government. This report further examined discrepancies by documenting differences in reading and writing abilities between children of varying ethnic backgrounds. For example, across fourth graders, African American and Hispanic students scored below the basic reading level at rates of 58% and 54%, respectively, as compared to only 24% of Caucasian students (NAEP, 2005).

Efforts to improve the literacy skills of all children and reduce the achievement gap have been the focus of many recent initiatives and laws, including the No Child Left Behind Act (NCLB), Response to Intervention (RtI), and School-Wide Positive Behavior Supports (i.e., SWPBS; Fuchs & Fuchs, 2006; NCLB, 2003; Sugai & Horner, 2002). While these programs have different focuses (overall school accountability, tiered levels of academic and/or behavioral support across multiple school settings, and behavioral changes), they all were developed to increase student achievement, particularly the achievement of children with the highest risk for academic failure. In addition to efforts at the elementary school level, early childhood proponents argue that education and intervention efforts prior to kindergarten entry are also important in supporting literacy outcomes for all children (Good, Gruba, & Kaminski, 2001; Torgeson, 2002; Walker & Shinn, 2002).

Evidence to support the need for early intervention as one of several efforts to improve literacy outcomes for all students comes from research demonstrating that identification of children at risk for poor reading outcomes can occur as early as kindergarten and first grade (Burke, Hagan-Burke, Kwok, & Parker, 2009; Cunningham & Stanovich, 1997; Juel, 1988; Scarbourough & Parker, 2003; Stanovich, 1986).

Students that exhibit deficits in pre-reading skills (e.g., letter identification or letter-sound correspondence) upon entry in kindergarten often struggle to progress in reading skills at a similar rate to their peers as they move through the elementary years (Burke et al., 2009). One study in particular showed that the rates of progress for students in first grade who were at the 50th percentile (middle) and the 10th percentile (low) were virtually indistinguishable through the first half of their first grade year. However, by the end of that first grade year, differences in reading trajectories had become apparent between these middle and low readers, and this "gap" continued to increase over time (Good, Simmons, & Smith, 1998). This frequently cited research identified that the problem for these "low" students was not a lack of progress entirely. The students in this "low" group were making progress, but were increasing their skills at a slower rate. In addition, approximately 84% of the students on the low trajectory were receiving either Chapter 1 or special education services. In other words, the students were receiving additional educational supports to remediate their performance deficits in reading; however, their low level of initial skills and slow rate of skill growth combined to make "catching up" nearly impossible (Good, Simmons, & Smith, 1998).

Many early intervention advocates have argued that differences in pre-reading skills upon entry to first grade suggest a need for early and intense intervention to prevent gaps in achievement (Burke et al., 2009; Burke, Hagan-Burke, Zou, Kwok, 2010; Snow et al., 1998; Torgeson, 2002). Research suggests that early intervention, or prevention, is more effective than interventions occurring at a later grade level (Burke et al., 2009; Cunningham & Stanovich, 1997; Juel, 1988; Scarbourough & Parker, 2003; Stanovich, 1986). Preschool programs are able to provide this early identification and early

intervention, and many argue that these early interventions should be used to close this achievement gap (Anderson, Shinn, Fullilove, Scrimshaw, Fielding, Normand, et al., 2003; Magnuson & Waldfogel, 2005; Ramey & Ramey, 2004).

Addressing Academic Performance Through Preschool Programming

A wide variety of preschool programs have been in existence for several decades including but not limited to programs such as Head Start, government-funded preschool programs intended for children indentified as "at-risk", private preschool programs, and day care centers that include "preschool-type" programming. In addition to federally funded programs, states began funding preschool initiatives as early as 1977. The mid 1990's marked a significant increase in the number of state-funded preschool programs, and state-funded preschool programs began to be tracked at a national level beginning in the new millennium (Barnett, Robin, Hustedt, & Schulman, 2003). During the 2001-2002 school year, forty-five state-funded programs were offered across the United States. Just under fifteen percent of the nation's four-year-old children were served by one of these programs. In the state of Wisconsin during the 2001-2002 school year, 166 school districts across the state offered four-year-old kindergarten (4K) programming. Nineteen percent of four-year-old children in the state of Wisconsin were enrolled in a state pre-kindergarten program during the 2001-02 school year, which translates to approximately 14,200 children.

While many of the initiatives described above focus on providing preschool for children of ethnic and/or a low socioeconomic family status, alternatively labeled 'at-risk', the most recent state-funded preschool initiatives have focused on offering educational options to all preschool-aged children whose parents wish to enroll them, and

are thus "universal". Universal preschool is becoming one of the leading efforts implemented by states to support the academic performance of all children. Proponents argue that universal preschools are more likely to be effective in reaching all targeted children, including low- and middle- income children, and when offered in a "universal" manner, may have larger and longer standing positive effects for children at-risk for poor educational performance (Barnett, Brown & Shore, 2004). Proponents also argue that, based on the results from program studies such as Head Start, the money spent on universal preschool programming is of greater value to overall educational and human services over time, with costs of universal preschool being significantly less than costs incurred later to intervene and further educate at-risk children and adults (Currie, 2000; Masse & Barnett, 2002; Temple & Reynolds, 2007; Zigler, Gilliam, & Jones, 2006).

Within the past decade, there has been an increase in preschool attendance as a result of state-funded preschool programs (Barnett, Carolan, Fitzgerald, & Squires, 2012). According to the 1970 US Census, 23 percent of four-year old children enrolled in a center or school-based preschool program. In 2001, the number of four-year old children in center or school-based preschool programs had risen to 66 percent (Magnuson, Meyers, Ruhm, & Waldefogel, 2004). According to the National Institute of Early Education Research's annual State of Preschool report for 2012 (Barnett et al., 2012) forty states offered state-funded preschool programs, eight of which fund universal preschool programs. In addition, four more states were working towards state-wide implementation of universal preschool programming at the time of the NIEER report (Barnett et al., 2012).

Statement of the Problem

As the types of preschool programming shift and the enrollment in state-funded universal programs increases, data on the effectiveness of universal preschool programming is needed. Program evaluations (such as Head Start, Perry Preschool, etc.) show a number of positive effects for early childhood education, but fewer studies have examined state-funded preschool programs. The existing program evaluations of state-funded preschools have demonstrated mixed results (Cambell & Ramey, 1994; Cambell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Reynolds, 2000; Reynolds, Temple, Ou, Robertson, Mersky, Topitzes, & Niles, 2007; Schweinhart, Barnes, & Weikart, 1993). Finally, few studies have examined the long-term effects of participation in universal preschool on overall reading achievement (Gormley & Gayer, 2003; Frede, Jung, Barnette, & Figueras, 2009).

This study examined the relationship between participation in a school district supported universal 4K program and subsequent literacy skills. The following questions were assessed:

1. Did students who participated in 4K enter Kindergarten with better-developed early literacy skills?
2. Is there a relationship between 4K participation and risk status for later reading difficulties?
3. Is 4K participation associated with higher reading performance on third grade reading assessments?

Definitions of Terms

Definitions for key terms used in the current study are provided below.

Four-year-old Kindergarten (4K): Unless stated otherwise, in the current study, 4K refers to universal, state-funded, four-year-old preschool programming.

Universal Preschool: Universal preschool makes access to preschool programming open to all families, without any financial restrictions or other eligibility requirements other than age and/or residence within the state or district providing the service. Universal preschool programs are voluntary to families, meaning that families can choose whether or not to participate in the program.

State-Funded Preschool: Preschool programming that is fully, or primarily, funded through state government funding is known as state-funded preschool. This is an alternative to privately-funded preschools that provide services through tuition paid for by families. State-funded preschools may also receive partial funding from local and federal government programs and grants. State-funded preschool is most commonly offered to children four years of age, though some programs also offer service to three year old children (Barnett, Hustedt, Robin, & Shulman, 2005).

Curriculum-Based Measurements (CBMs): CBMs, specifically in the area of reading, are measures of skills that are defined as being standardized, short in duration, and given several times throughout an instructional year. Pre-determined goals are designated; analyses are completed to determine the level of performance or proficiency needed to be on track for attaining future literacy goals, such as state-mandated test scores or graduation standards. These predetermined goals, or benchmarks, are then used as comparison to individual student performance. Students not performing to benchmark are identified as "at-risk" for future reading difficulties, and should receive additional intervention and support. At the early stages of literacy knowledge, CBMs are most often

composed of letter or letter sound identification and/or identification of phonemes. Most prevalent reading CBMs consist of oral reading fluency probes, which require students to read aloud from a leveled passage for one minute; words correct per minute are gathered and used to compare to benchmark goals.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS): DIBELS (Good & Kaminski, 2002) is a curriculum-based measure specifically designed to assess early literacy skills that are precursors to fluent reading. Skills assessed include phonological awareness, alphabetic principle, and reading accuracy and fluency (Good, Gruba, & Kaminski, 2001; Good & Kaminski, 2002).

CHAPTER II

Review of the Literature

This chapter briefly reviews research on the effectiveness of preschool programming. The history of state-funded preschool programs, approaches to early intervention preschool programs, and current research on the efficacy of preschool programs is reviewed. In addition, a critical analysis of the research is provided.

Effectiveness of Early Intervention

Efforts to provide early intervention to children at risk for learning challenges are long standing. Head Start is one of the most widely recognized readiness programs in the US. Started in 1965, it is a program of the United States Department of Health and Human Services, which aims to provide "comprehensive education, health, nutrition, and parent involvement services to low-income children and their families" (U.S. Department of Health and Human Services, Administration for Children and Families (US DHHS ACF), 2005). Head Start has been researched thoroughly for decades, and has demonstrated multiple positive findings for cognitive skills, academic skills, social-emotional skills, child heath, and parenting skills (US DHHS ACF, 2005). According to the fact sheet report of 2012, over one million children ages birth to five years were enrolled in Head Start programming, with 48% of participants being four years old (U.S. DHHS ACF Head Start, "Program Information Report", 2013). Children receiving at least one year of Head Start programming were found, at the end of their Head Start year, to demonstrate significant stronger cognitive skills, specifically vocabulary, letter-word identification, spelling, color identification, letter naming, and pre-academic skills (US DHHS ACF, 2005).

Other early intervention programs, such as the High/Scope Perry Preschool Project (Schweinhart et al., 2005), the Carolina Abecedarian program (Campbell et al., 2002), and Chicago's Child-Parent Center (Reynolds et al., 2007) have also documented positive effects on children's cognitive and academic skills from the intensive early intervention provided by each program. All four programs have demonstrated that program participants show immediate positive outcomes for academic achievement measures in the areas of reading and math, as well as cognitive or intelligence measures. In addition, long-term benefits are documented for participants in all four programs, such as higher levels of educational achievement and lower levels of criminal behaviors and arrests. The Chicago Child-Parent Center and the Abecedarian Project programs have also documented significantly higher levels of occupational status, as measured by higher-paying jobs and higher-skill jobs, higher scores on intellectual ability measures through childhood, and lower rates of social service needs (such as welfare, out-of-home child placements, and teen pregnancy) for participants (Campbell & Ramey, 1994; Campbell et al., 2002; Reynolds, 2000; Reynolds et al., 2007; Schweinhart, Barnes et al., 1993).

Recently, Puma et al. (2010) specifically examined the impact of Head Start participation on four year old students. Literacy and pre-academic skills for four year olds participating in Head Start were compared across two groups. Group one included children who attended a Head Start program site, and group two included children who were eligible but did not attend a Head Start program site (i.e., may have attended an alternative care setting, or may have been primarily in parental care). Head Start participants (group one) demonstrated stronger vocabulary, letter and word identification,

spelling, 'pre-academic skills', color identification, and letter naming when compared to similar age and SES level children. In addition, parent reports, through the use of the Emergent Literacy Scale, also indicated stronger literacy skills for participants versus non-participants (Puma et al., 2010).

However, it is important to note that these intensive programs, while demonstrating positive effects, have specific regulations for service delivery and are subject to stringent entrance criteria. All four programs primarily serve children from homes or communities of differing ethnicities or races, and have strict entrance criteria for family income levels. For example, Head Start serves children of low socio-economic status and requires a significant level of parent participation as determined by site-based regulations. The High/Scope Perry program was designed exclusively for African American children; the Child-Parent Center focused on African American and Hispanic children from low-income families; and the Abecedarian Project primarily served African American children from low income and single-parent homes. Therefore, while these programs demonstrate findings to support the positive effects of early intervention, they differ significantly from universal preschool programs in many ways.

The Child-Parent Center and High/Scope Perry programs required significant parent involvement, including in-home interviews, parent involvement within the school setting, and follow up surveys and interviews. Head Start programs had varying levels of eligibility and parent involvement requirements, which were determined by each site, and generally offered 24 to 28 hours of care per week. The Abecedarian Project program provided educational intervention in both preschool and primary years (eight years total), in preschool only (five years total), or during primary grades only (three years total), and

delivered intensive interventions ranging six to eight hours per day for five days a week and 50 weeks per year. While these strict criteria for participation are appropriate for identifying children with the greatest risks for poor school performance, the programs do not allow all children or all potentially at-risk children to receive early education programming.

In contrast to the strict criteria for Head Start, Child-Parent Center, High/Scope Perry, and Abecedarian, many state-funded preschools have fewer entrance criteria. National trends from 2005 indicate that over 75% of preschool programs had an income requirement in addition to child age and area of residence (Barnett, Hustedt, Robin, & Shulman). In addition to differences between entrance criteria between Head Start, High/Scope Perry, Child-Parent Center, and Abecedarian and state-funded preschools, state-funded programs may also vary from one another in terms of entrance criteria, the amount of schooling or care offered per day or per week, teacher licensure requirements, and the amount of parent outreach and education provided (Early, Howes, Kraft-Sayre, Barnett et. al., 2005). In further contrast, few universal preschool programs have eligibility requirements, other than age and area residence. Many universal preschool programs follow a similar calendar to the local school district (such as summers and holidays off), and may offer programs for half days or for school hours (Barnett, Hustedt, Robin, & Schulman, 2005).

Because of these wide differences between and among programs, educators need to realize that the positive outcomes documented for these established programs (such as High/Scope Perry, Abecedarian, etc.), while promising, cannot be directly applied or generalized to other early intervention programs, particularly state-funded or universal

preschool programs. This points to a greater need for closer inspection of state-funded programs, particularly universal programs, due to the potential for widely varying populations between children accessing these programs.

The Effectiveness of State-Funded Preschool Programs

Gilliam & Zigler (2001) completed a critical meta-analysis of all impact evaluations of state-funded preschools from 1977 to 1998. The meta-analysis involved existing state-funded preschool programs that were specifically designed for children from homes identified with a low socioeconomic status. Of the 13 studies reviewed, ten used comparison groups to determine program effects. All ten evaluations varied in terms of number of participants, number of cohorts evaluated, areas of outcomes assessed, and ages at which measures were completed.

According to Gilliam and Zigler (2001), positive outcomes were observed in seven of the ten studies. Seven studies used achievement tests, such as the California Achievement Test or the Iowa Test of Basic Skills, to measure reading and math skills for program participants once they were enrolled in kindergarten through third grade. Evaluations from the states of Florida, Maryland, South Carolina, and Texas all documented significant differences between participants and non-participant in both reading and math. Results from Florida indicated an effect size of .23 in the area of reading achievement for measures administered during kindergarten. Several states also demonstrated lasting differences between participants and nonparticipants in subsequent grades. Georgia and South Carolina reported higher achievement scores for program participants in first grade (effect sizes of .24 and .07 respectively) and Maryland and Texas reported higher achievement scores for preschool participants in third grade (.39

and.08 effect sizes, respectively). Authors of the meta-analysis highlighted the importance of high quality evaluations for accountability and monitoring of state-funded preschool programs, stating that in order to be able to draw meaningful conclusions from these program evaluation studies, appropriate research methodology must be employed (Gilliam & Zigler, 2001).

Gilliam & Zigler (2001) identified significant concerns for the research design of the studies included in their review of evaluations from 1977 to 1998. Concerns were noted for poor samples; sample sizes ranged significantly from 22 pairs of matched students, which resulted in weak statistical power, to several thousand preschool participants and several thousand comparison peers, which significantly impacted the researchers' ability to reject the null hypothesis. Concerns were also indicated for poor measurements. A wide variety of assessments were used in the ten primary evaluations analyzed, and while several of them are well-known and psychometrically valid instruments, many instruments used were "relatively unknown… with little data regarding their reliability and validity" (Gilliam & Zigler, 2001). Finally, concerns were noted for poor study methodology. Only 10 of the total studies reviewed used a contrast group to compare preschool participants; of those ten, only a few worked to control for baseline differences by either matching the at-risk status for each group or by using pretests for both groups. Very few studies calculated effect sizes, which limited the generalization and understanding the findings of each evaluation.

Starting in 2005, the National Institute of Early Education Research (NIEER) began reporting on the effectiveness of state-funded preschool programs (Frede, Jung, Barnett, & Figueras, 2009; Frede, Jung, Barnett, Lamy, & Figueras, 2007; Hustedt,

Barnett, & Jung, 2008; Lamy, Barnett, & Jung, 2005a; Lamy, Barnett, & Jung, 2005b; Lamy, Barnett, & Jung, 2005c; Lamy, Barnett, & Jung, 2005d; Wong, Cook, Barnett, & Jung, 2007).

The NIEER report also highlighted another related methodological concern for program evaluation studies of state-funded preschool programs, specifically, selection bias. Selection bias is a concern when evaluating state-funded preschool programs because "Children who attend state preschool programs differ from those who do not … because the programs select the children, and the families select the programs" (Lamy, Barnett, & Jung, 2005c). For example, the town or neighborhood in which the child lives may or may not offer a state-funded preschool program, the programs may or may not have entrance criteria for participation, or the family may have access to alternate preschool experiences (such as day care centers, in-home care, or faith-based preschool) that the family prefers.

To minimize the potential effects of selection bias, the NIEER studies applied a Regression-Discontinuity Design (RDD) model in the evaluation of outcomes for state-funded preschool participation (Imbens & Lemieux, 2008; Thistlethwaite & Campbell, 1960). The RDD approach relies on a strict adherence to the idea that eligibility for preschool (and therefore kindergarten) is determined by date of birth alone. This rule allows the construction of two groups, one entering kindergarten that has already attended a preschool program at age four, and one just entering a preschool program at age four, and thus have not received the "treatment" of preschool yet. The RDD approach can be likened to a randomized trial for all children near the preschool program's enrollment cutoff age. The RDD creates a sample of children of similar age,

some who were born a few days before the enrollment date cutoff and others were born a few days later (Frede, Jung, Barnett, & Figueras, 2009). By testing all of the children at the same time, regardless of "treatment group" status, the authors argue that they are able to obtain an unbiased estimate of the preschool program's effectiveness (Frede, Jung, Barnett, & Figueras, 2009).

Selection criteria for participation in preschool programming was a particular area of interest to the current study, and thus, the following review of programs will be divided into those with selective criteria for participation and those which are universal, or available to all residents. State-funded preschool programs from six states were reviewed in the NIEER study, with two states having two studies each. Of the six states examined, there was a mix of programs with selective entrance criteria versus universal programming. Two states offered preschool programming to students from low-income families and at least one other risk factor (e.g., dual language, single parent home, special education need, etc.). Two states offered universal access to children living within specific school districts or sites that were identified as being at high risk for poor academic performance due to lower performance indicators at the elementary schools. One state identified that most of its counties offered universal programming with a small number of counties only offering programming to at-risk children. One state focused only on the state's larger cities, where universal programming was offered across the entire city. Finally, three of the six states offered programming to four-year-old children exclusively, while the other three states also offered programming to three-year-old children who had been identified as either needing special education or as high risk.

In addition to entrance criteria, programs within each state also varied in terms of length of programming. Five of the six states offered an average equivalent of one half-day programming for 5 days per week. Other states indicated that hours of operation were determined locally. All six states identified they offered preschool programs during the academic year only. Despite the variability in program details, positive findings were reported across several measures of academic skills. Specifically, small effects were reported for language development, small to medium effects were reported for math, and medium to large effects were reported for print awareness (Barnett, Lamy, & Jung, 2005). The eight NIEER studies are reviewed below.

In their program evaluations, the states of Michigan, New Mexico, South Carolina, West Virginia and Oklahoma all incorporated measures of receptive vocabulary using the Peabody Picture Vocabulary Test – Third Edition (Dunn & Dunn, 1997). Studies from Michigan, South Carolina, New Jersey, West Virginia, and Oklahoma also included measures of early literacy skills, specifically Sound Blending and Print Awareness from the Preschool Comprehensive Tests of Phonological and Print Processing (Pre-CTOPPP; Lonigan, Wagner, Torgesen, & Rashotte, 2002). Sound Blending consists of the child using words to create compound words (horse… shoe… horseshoe) or word parts (i.e., phonemes) to create words (ca…p… cap). Print Awareness consists of a variety of skills, including knowledge of letters and sounds and identifying the parts of a book (e.g., title, reading left to right, etc.). The New Mexico study included the Test of Preschool Early Literacy (Lonigan, Wagner, Torgesen, & Rashotte, 2007), which is the published and standardized version of the Pre-CTOPPP. Finally, several studies assessed childrens' math skills using the Applied Problems

subtest of the Woodcock-Johnson Tests of Achievement – Third Edition, which assesses the student's ability to solve story-type math problems (Woodcock, McGrew, & Mather, 2001).

It should be noted that across all NIEER studies using the RDD design, student scores, and therefore effect sizes, were calculated using raw scores or percentage of items answered correctly. Therefore, information about the relative performance of each group performed in relation to a standardization sample was not reported. This method does not allow for a determination of how the overall performance relates, that is, whether their overall score is in the average range or not. Rather, this method demonstrates whether participants differed from non-participants in the areas assessed (Frede, Jung, Barnett, & Figueras, 2009; Frede, Jung, Barnett, Lamy et al., 2007; Hustedt et al., 2008; Lamy, Barnett, & Jung, 2005a; Lamy, Barnett, & Jung, 2005b; Lamy, Barnett, & Jung, 2005c; Lamy, Barnett, & Jung, 2005d; Wong et al., 2007).

The Michigan School Readiness Program targeted at-risk four-year-olds, identified by income and/or other risk factors. Program evaluation assessments were completed upon entry to the kindergarten/preschool academic year, using the RDD design, to compare children who had completed the 4K program to those who had not. When compared to children who had not attended the preschool program, preschool participants demonstrated significantly higher skills on assessments of receptive language, math, and print awareness. Assessments of phonological awareness (specifically the skill of blending) did not show differences of statistical significance, as measured by the Pre-CTOPPP (Lamy, Barnett, & Jung, 2005a). The effect sizes, as

reported by Lamy, Barnett, & Jung (2005a), range from 0.21 in receptive language, to 0.44 in math, and .96 in print awareness.

South Carolina developed a Half-Day Child Development Program. Programs were developed at a local level, and the 'at-risk' factors used for eligibility were determined locally. Assessments were also completed upon entry to the kindergarten/preschool academic year following an RDD design. Students who attended the South Carolina state-funded preschool programs demonstrated significantly higher scores in the areas of receptive language and print awareness (Lamy, Barnett, & Jung, 2005b) with effect sizes of 0.35 for receptive language and 0.71 for print awareness. Similar to the Michigan study, measures of phonological awareness skills did not achieve differences of statistical significance. Measures of math were not assessed.

New Mexico's state-funded PreK program focused on students living within the most at-risk neighborhoods in the state, which were identified through the use of several criteria outlined by the state department of education. Access to programming was universal for four-year-old children living within these neighborhoods. Based on standardized measures completed upon entry of their kindergarten/preschool year, using the RDD design, students attending the New Mexico PreK programs demonstrated higher scores on assessments of receptive language, math, and print awareness as compared to students in the same neighborhood who had not attended the program (Hustedt et al., 2008). Effect sizes, as reported by Hustedt et al. (2008), include 0.25 for receptive language, 0.5 for math, and 0.59 for print awareness.

The Abbott Preschool Program of New Jersey, initiated in the 1999-2000 school year, provided full day preschool for three- and four-year-old children living within

specified low-income school district boundaries. In addition, wraparound services (before- and after- school care and summer programs) were also provided. An initial report in 2007 demonstrated that when compared to nonparticipants, children who had attended the Abbott Preschool program had significantly stronger skills on measures of oral language, early literacy skills, and math completed at the beginning of their kindergarten year (Frede, Jung, Barnette, Lamy et al., 2007). Notably, the differences between groups in each of these skill areas was approximately two times more for the children who attended two years of the program (for age three and age four) versus children who attended one year (for age four only).

When West Virginia's preschool program was evaluated in 2005, Lamy, Barnett, and Jung (2005c) indicated that in many counties, all children were eligible for 4K as long as they met the age requirements and lived in participating school districts. Only a small number of counties determined program eligibility based on specific risk status factors, like family income. Therefore the evaluators of the program described West Virginia's program as universal. West Virginia's Early Education Program demonstrated results similar to other states. Assessments completed at the beginning of the kindergarten/preschool year, using the RDD design, indicated that students attending the state-funded programs demonstrated stronger performance on measures of receptive language, math, and print awareness as compared to peers who did not attend these programs. Assessments of phonological skills were not significantly different when participants were compared with non-participants. Effect sizes, as reported by Lamy, Barnett, and Jung (2005c), indicate 0.27 for receptive vocabulary, 0.41 for math, and 0.93 for print awareness.

Oklahoma, specifically the Tulsa Pre-Kindergarten program, provides universal preschool programs to all four-year-old children within Tulsa school boundaries. Program evaluation findings were similar to other states. Preschool participants entered kindergarten with significantly higher vocabulary, math and print awareness skills, but differences in phonological awareness were not noted (Lamy, Barnett, and Jung, 2005d). An earlier evaluation of Tulsa Public School's Pre-K Program (Gormley & Gayer, 2003) was consistent with the NIEER report. In the Gormley & Gayer (2003) study, teachers completed the Early Childhood Skills Inventory for preschool participants and non-participants. Compared to their peers who did not participate, participants were rated as demonstrating significantly higher academic skills in the areas of language, cognitive/knowledge, and motor skills. The largest effects were observed in language. In addition, students of low socioeconomic status or of a racial/cultural minority group demonstrated the greatest benefits in these three areas, with Hispanic children showing the greatest effects. Notably, differences in social-emotional development, as measured by this inventory, were not noted.

The APPLES Blossom, an interim report published by NIEER in 2009, employed a traditional longitudinal analysis to evaluate the academic progress of Abbott participants after kindergarten. As with the initial study (Frede, Jung, Barnett, Lamy et al., 2007), greater effect sizes for academic outcomes were noted for children who attended two years of the Abbott program (at ages three and four) as opposed to only one year (only age four). In this 2009 follow up study, participants in the Abbott preschool program outperformed children who did not attend the program in the areas of oral language and math at the end of their kindergarten, first grade, and second grade years.

However, reading skills did not demonstrate the same pattern. First and second grade measures included assessments of word reading, reading fluency, and passage comprehension from the Woodcock-Johnson Tests of Achievement – Third Edition, and the difference in scores between participants and nonparticipants did not reach statistical significance (Frede, Jung, Barnett, & Figueras, 2009).

Summary and Critical Analysis of the Research

Overall, the results of these studies suggest that state-funded preschool programs do demonstrate positive effects for children who participate, especially in the areas of language, math, and print awareness. NIEER cautions against directly comparing results from state to state because of state differences on variables such as age of participants, length of program, and program accessibility. Comparisons of effect sizes enforce this caution. The data suggests that there is no clear pattern for greater or lesser effects in targeted programs versus universal programs. For example, scores range between .21 (i.e., Michigan, targeted children) to .35 (i.e., South Carolina, targeted children) for receptive vocabulary, between .29 (i.e., Oklahoma, universal) to .5 (i.e., New Mexico, universal by neighborhood) for math skills, and between .56 (i.e., New Jersey, targeted by neighborhood) to .96 (i.e., Michigan, targeted children) for print awareness. The significant difference between program effects further strengthens the argument that individual evaluation for state-funded preschool programs is needed.

In addition to positive effects on pre-academic skills assessed at kindergarten entry, findings from preschool program evaluations in New Jersey (Frede, Jung, Barnette, & Figueras, 2009) suggest that more is better. Children who attended two years of pre-school, versus those who attended only one year, demonstrated significantly stronger

academic skills (Frede, Jung, Barnett, & Figueras, 2009). Further, the Abbott study demonstrated that children who participated in two years of Abbott preschool demonstrated significantly stronger receptive language and math skills in second grade.

Despite the promising results reported in the state-funded preschool studies reviewed, limitations of the existing program evaluations support a need for further research and different approaches to preschool program evaluations. First, few states have evaluated longitudinal data for preschool participants. Recent studies have focused on the differences of children at the beginning of, or during their, kindergarten year (Frede, Jung, Barnett, Lamy et al., 2007; Gormley & Gayer, 2003; Husted et al., 2008; Lamy, Barnett, & Jung, 2005a; Lamy, Barnett, & Jung, 2005b; Lamy, Barnett, & Jung, 2005c; Lamy, Barnett, & Jung, 2005d; Wong et al., 2007). This information, while valuable, only demonstrates that children are starting out on a more even playing field, and not necessarily whether those advantages are maintained. As early intervention programs, such as state-funded preschool, are being used to argue for closing the achievement gap, it becomes more important to examine the lasting effects of these programs.

Further, the NIEER findings emphasize the positive impact that high-quality state-funded preschools can have on pre-academic skills for children entering kindergarten, specifically receptive language, math skills, and print awareness. However, several other pre-academic and social skills are important for children to experience success in kindergarten and later grades. Social skills, behavioral and emotional functioning, self-help skills, and motor skills are also related to overall student achievement in elementary school (Howse, Lange, Farran, & Boyles, 2003; Huffman,

Mehlinger, & Kerivan, 2000; National Institute of Child Health and Development Early Childcare Research Network, 2003). Only one of the aforementioned studies examined the relationship between preschool participation and social skill development (Gormley & Gayer, 2003) and findings noted no significant differences between participant and nonparticipants on teacher ratings. More research is needed to evaluate the effects of preschool participation on the behavioral-readiness of children entering kindergarten.

Assessments used to measure academic performance were similar in most of the studies reviewed; namely, norm-referenced standardized assessments of academic and pre-academic skills. Evaluations of Head Start, the High/Scope Perry Preschool study, the Abecedarian Project, and the Chicago Child-Parent Center employed similar academic measures. While norm-referenced assessments play an important role in identifying student skills relative to similarly aged children, scores from these types of assessment do not provide an indication of differences in risk-status for future academic failure. Stated differently, norm-referenced measures are designed to assess a student's current level of performance relative to other children of the same age, but they do not always allow for the determination of whether a student is demonstrating adequate mastery of skills pre-requisite to the academic demands of kindergarten and later grades. Because each of the studies reviewed compared mean scores of preschool participants and non-participants on norm-referenced measures, the findings do not directly answer the question of whether participating in a preschool program was associated with differences in student risk for academic failure upon kindergarten entry, which is an important part of identifying at risk students early and closing the achievement gap.

The more recent NIEER studies (Frede, Jung, Barnett, Lamy et al., 2007; Hustedt et al., 2008; Lamy, Barnett, & Jung, 2005a; Lamy, Barnett, & Jung, 2005b; Lamy, Barnett, & Jung, 2005c; Lamy, Barnett, & Jung, 2005d; Wong et al., 2007) attempted to address many of the concerns noted within the meta-analysis with the RDD method of creating a matched control group. While this method of subject selection can address selection bias more effectively, RDD has drawbacks. Namely, the RDD methodology cannot be used within longitudinal designs. Therefore, the RDD allows for greater confidence in interpretations of between group differences in skills at a specific point in time, but it does not allow for an evaluation of long term differences in academic performance over time. Additional longitudinal evaluations are needed to determine whether differences in pre-academic skills upon kindergarten entry are maintained and whether preschool participation is associated with enhanced academic performance beyond kindergarten.

Current Study

As educators have worked to determine accurate methods of identify struggling students for early academic intervention, research has determined that developmental reading trajectories can be identified and examined using curriculum based measurements (CBM) and that CBM reading measures can provide valid and reliable measures of overall reading proficiency, as well as strong predictors of future reading skills (Shinn, Good, Knutson, Tilly, & Collins, 1992). Numerous studies of reading CBMs have identified differences in performance between successful and struggling readers and differences in the trajectory of reading growth for the same groups of students (Good, Simmons, & Smith, 1998). CBMs, such as the Dynamic Indicators of Basic Early

Literacy Skills (DIBELS), can be used both as an academic screener to identify students at risk for poor reading performance and as a progress monitoring tool to track student growth in reading performance (Kaminski, Cummings, Powell-Smith, and Good, 2008). CBMs are developed to be used over a longer period of time, and thus are able to identify long-term differences in academic performance by each student.

The present study evaluated the effects of participation in a four-year-old kindergarten program (4K) on early literacy skills and reading development. Student performance on early literacy CBMs and state-mandated reading assessments was evaluated to determine 1) whether a relationship existed between a child's participation in 4K and their performance on early literacy measures completed at the beginning of kindergarten, 2) whether participation in 4K was related to differences in reading risk status as determined by assessments in kindergarten, and 3) whether differences in third grade reading performance existed between 4K participants and non-participants.

Chapter III

Methodology

This study evaluated literacy outcomes for students participating in a school district sponsored four-year-old kindergarten program (4K) located in the state of Wisconsin. Early literacy assessments in kindergarten and state-mandated reading assessments in third grade were used to compare the literacy skills of 4K participants and non-participants and to evaluate the relationship between 4K participation and literacy development. Existing data from the school district was used to address the research questions. The selection of participants, assessment tools, and procedures for data collection are described below, followed by a discussion of data analysis procedures.

Participants

At the time of data collection, a universal 4K program had been in place for four years in the collaborating school district. In the school district, 4K is offered free of charge to all families residing within the district geographic boundaries. Children must be four years of age by September 1st to enroll in the 4K program. The district provides educational experience to approximately 11,000 students in grades 4K to twelfth grade. Although there are twelve sites serving elementary-aged students, only five elementary schools were included in the current study. The cooperating school district for the current study had pre-established procedures to complete Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) assessments at the elementary schools with the highest rates of free/reduced lunch eligibility. Therefore, assessment data for students only enrolled in these five schools were analyzed for this study. The

percentage of students eligible for free or reduced lunch ranged from 37 percent to 77 percent across the schools.

The data collected included five cohorts of students (the last cohort with No 4K option, the first 4K class year, the second 4K class year, the third 4K class year, and the fourth 4K class year) who completed DIBELS assessments during their kindergarten year. At the time of data collection, two of the cohorts had completed third grade. Due to an overrepresentation of Special Education and English Language Learner students, data from these two populations were removed. Tables 1 and 2 outline the demographic information for participants, as gathered from the district's student information system. The cooperating district could not provide individual information as to free and reduced lunch status by student; therefore precise percentages for this variable could not be reported.

Enrollment in the 4K program increased each year, with 37 % of kindergarteners having had the 4K experience during the first cohort year and 68 % of kindergarteners having the 4K experience just three years later. Gender rates were stable for all groups. Similarly, ethnicity rates were also consistent across the five cohorts. White children made up the majority of all cohorts (approximately 70-80 %) with Asian ethnicity being the next largest racial group (approximately 9-11 %). The number of children who were identified as eligible for English Language Learner services or as eligible for Special Education services during their kindergarten year or later decreased from the first year reported to the fifth year reported. The number of children being retained or advanced during kindergarten or any later grades also decreased from the first year to the fifth year.

Table 1

Participant Demographics

Cohort	4K Participants (%)	Male (%)	Ethnicity (%)				
			White	American Indian	Asian	African-American	Hispanic
No 4K (N=450)	—	232 (51.5)	76.0	1.3	10.4	5.7	2.8
1st 4K (n=401)	148 (36.9)	206 (51.3)	76.8	2.5	11.4	4.4	3.7
2nd 4K (n=386)	163 (50.5)	185 (47.9)	78.0	1.0	10.8	8.8	1.2
3rd 4K (n=324)	164 (50.6)	170 (52.4)	84.2	0.6	9.2	3.3	1.8
4th 4K (n=271)	185 (68.2)	139 (51.3)	83.0	1.4	9.5	4.4	1.4

Table 2

Participant Special Services and Attrition by Cohort

Cohort	Identified during Kindergarten or later			
	ELL	SPED	Retained	Advanced
No 4K	43	86	1	1
1st 4K	34	61	6	0
2nd 4K	27	53	1	2
3rd 4K	21	33	0	0
4th 4K	3	13	0	0

Procedure

Following approval for the study by the university's Institutional Review Board (approval number 2010-2011-060), existing academic assessment data, specifically DIBELS benchmark assessment scores and third grade Wisconsin Knowledge and Concepts Examination (WKCE) reading scores; data on preschool program participation status; and demographic information including gender, ethnicity, grade level, special education eligibility, and eligibility for English Language Learner programming was

obtained from the school district. All data were gathered from the school district's student information system for students enrolled in the district between the 2004-2005 and 2008-2009 academic years. District assigned student identification numbers were used to match kindergarten assessment data with third grade assessment data. All other unique identifiers were omitted to protect student privacy. Because the archival data was owned and collected by the school district, parent permission was not required. The participating school district received results of the program evaluation for their own use; otherwise, no incentives were provided for the school district to release student information.

The 4K program offered student contact for half days (i.e., morning or afternoon) four days per week, following the school district calendar for holidays and non-student contact days. Classrooms were offered at either district sites (such as within one of the elementary school buildings) or out in a community site (such as at a specified day care center). 4K teachers were hired by the school district, and met all requirements for teaching licensure at that age level. Specific information as to which site each 4K participant attended was not available, and therefore was not controlled for.

Measures

Dynamic Indicators of Basic Early Literacy Skills (DIBELS). DIBELS assessments are part of a system of curriculum-based measurements designed to assess early literacy skills. The measures are fluency-based and are designed to predict future reading performance by targeting early literacy skills found to be predictive of later reading success (Good & Kaminski, 2002). DIBELS subtests were designed to reflect the early literacy skills emphasized in the National Reading Panel report (NICHD, 2000),

including skills such as phonemic awareness, phonics/alphabetic principle, reading fluency, and comprehension. According to the authors, DIBELS can be used for screening purposes, to identify students "at risk" or "in need of intervention", to monitor overall student progress through the practice of tri-annual benchmark assessments, or for more frequent progress monitoring on an individual level. The early literacy measures have been found to be predictive of future reading skill performance (Dynamic Measurement Group, 2008; Goffreda & DiPerna, 2010; Good & Kaminski, 2002; Wayman, Wallace, Wiley, Ticha, & Espin, 2007). The fall and winter benchmark assessments for kindergarten were included in the current study, and are discussed below.

Initial Sound Fluency (ISF). ISF measure is designed for students in the beginning and middle of the kindergarten year. The ISF task is a measure of phonological awareness and assesses the student's ability to identify beginning sounds in words. For ISF, students are asked to identify one picture out of four that represents a word matching the initial letter sound presented orally by the examiner. For example, a student is shown four pictures: mouse, flowers, pillow, and letters. The examiner asks the child to point to the picture that begins with the /fl/ sound (flowers), the /l/ sound (letters) and the /p/ sound (pillow). The task is timed to determine how many initial sounds a student can identify in one minute. Every fourth question asks the student to produce the beginning sound for a specific picture. For example, the examiner points to a picture of a mouse and asks the student what sound it begins with.

According to the instructional recommendations within the administration and scoring manual (Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002, as cited in Good & Kaminski, 2002), students scoring at or above 8 initial sounds per minute at the

fall kindergarten benchmark are considered "low risk" for future reading difficulties, and those scoring at or above 25 initial sounds per minute at the winter benchmark are considered established in this skill, and thus the 'odds are in favor of achieving subsequent outcomes' (Good, Simmons, Kame'enui et al., 2002). According to the DIBELS Data System Technical Report 1102 (Cummings, Otterstedt, Kennedy, Baker, & Kame'enui, 2011), the mean score on the ISF measure is 11.25 (9.69 SD) at the fall benchmark and 27.42 (16.57 SD) at the winter benchmark.

According to the DIBELS technical manual (Dynamic Measurement Group, 2008), measures of ISF reliability range from .61 to .95. ISF demonstrates high to medium concurrent validity when compared to the Comprehensive Test of Phonological Processing (r = .46-.60), the Woodcock-Johnson reading measures (r = .36), the Phoneme Segmentation Fluency task from DIBELS (r = .51), and the Nonsense Word Fluency task from DIBELS (r = .51). Predictive validity measures indicate adequate to strong correlations, ranging from .29 (Nonsense Word Fluency task of DIBELS) to .46 (Woodcock-Reading Mastery Test-Revised), when measures were administered in the winter of the kindergarten year. According to Good, Kaminski, Shinn, et al. (2004), the median predictive validity of kindergarten ISF is .38 as compared to the DIBELS ORF in spring of first grade, and .36 as compared to the Woodcock-Johnson Psycho-Educational Battery total reading cluster standard score.

Letter Naming Fluency (LNF). LNF is administered to students throughout the kindergarten year and in the fall of the first grade year. LNF measures a student's letter knowledge, specifically how many upper or lower case letters a student can identify in one minute. The student is presented with a page of upper and lower case letters, which

are presented in random order. Students are informed that they will be told the letter if they cannot identify it. During the exercise, if a student pauses at a letter for three seconds, the examiner provides the letter name and the student is encouraged to continue. According to the instructional recommendations within the administration and scoring manual (Good, Simmons, Kame'enui et al., 2002, as cited in Good & Kaminski, 2002), students scoring at or above 8 letter names per minute during the fall kindergarten benchmark, and 27 or more letter names at the winter kindergarten benchmark, are considered "low risk". Conversely, students who score 2 or fewer letter names at the fall of kindergarten benchmark and 15 or fewer letter names at the winter of kindergarten benchmark are considered "at risk" for future reading problems.

According to the DIBELS Data System Technical Report 1102 (Cummings, et al., 2011), the mean score for LNF at the fall benchmark is 17.19 letter names per minute (15.26 SD). At the winter benchmark for kindergarten, the mean score on the LNF task is 37.44 (17.19 SD).

According to the DIBELS technical manual (Dynamic Measurement Group, 2008), reliability measures range from .89 to .98 for LSF. Concurrent validity results ranged from low to high, with the strongest correlation (.70) to reading tasks of the Woodcock-Johnson and the weakest correlation (.24) to the Word Usage Fluency task from DIBELS (Dynamic Measurement Group, 2008). Predictive validity measures indicate strong correlations, ranging from .42 (Terra Nova reading tasks, as completed during the spring of first grade) to .73 (TOWRE, as completed during the middle of first grade). LNF, or naming speed, tasks are found to be highly predictive of future reading performance, and thus, are widely used CBMs (Kirby, Parrila, & Pfiffer, 2003; NICHD,

2000; Speece, Mills, Ritchey, & Hillman, 2003). Rouse and Fantuzzo (2006) was able to demonstrate a "unique predictive relationship between early letter naming ability and later reading competencies while concurrently controlling for the variance associated with phonological skills and print conventions". Simultaneous regression analyses completed showed LNF to be the most significant predictor of first grade reading, vocabulary, and language (Rouse & Fantuzzo, 2006).

Nonsense Word Fluency (NWF). Nonsense Word Fluency (NWF) is administered at the middle of the kindergarten year through to the beginning of the second grade year. This assessment is a measure of the alphabetic principle and specifically measures that child's knowledge and application of the letter-sound correspondence and the ability to blend letters sounds to read words. The student is presented with nonsense words, constructed of vowel-consonant or consonant-vowel-consonant combinations, and asked to orally produce the letter sound of each letter, or orally produce the whole nonsense word. A measure is taken of how many sounds the student is able to produce in one minute, with the intent being that the student is able to fluently "read" the letter sounds as whole words. According to the instructional recommendations within the administration and scoring manual (Good, Simmons, Kame'enui et al., 2002, as cited in Good & Kaminski, 2002), students producing 13 or more correct letter sounds per minute on the winter kindergarten benchmark are in the "low risk" range, while students producing 5 or fewer correct letter sounds per minute are in the "at-risk" range. According to the DIBELS Data System Technical Report 1102 (Cummings, et al., 2011), the mean score for NWF during the winter of kindergarten is 24.80 Correct Letter Sequences (18.59 SD).

According to the DIBELS technical manual (Dynamic Measurement Group, 2008), NWF demonstrates strong reliability, ranging from .86 to .98. Concurrent validity measures indicate variable results, with the strongest correlations to reading tasks of the Woodcock-Johnson Revised (r = .91) and the weakest correlation to the TPRI Early Reading Assessment (r = .27). Predictive validity measures indicate very strong correlations, with scores ranging from .55 (TerraNova reading tasks, as measured during the spring of first grade) to .77 (Nonsense Word Fluency during the spring of first grade).

Phoneme Segmentation Fluency (PSF). Phoneme Segmentation Fluency (PSF) is administered beginning in the middle of the Kindergarten year and throughout the entire first grade year. This measure assesses a student's ability to segment, or take apart, a word into the individual phonemes. The examiner presents a word with three or four phonemes (such as "sat" or "land") and the student is asked to verbally reply with the individual phonemes (/s/…/a/…/t/ or /l/…/a/…/n/…/d/). Scores are based on how many correct phonemes the student produces within one minute. If the student does not segment at an individual phoneme level (such as response of /m/…/at/), the student would earn points for the phonemes given, such as two points for /m/…/at/, rather than the three points for /m/…/a/…/t/ (Good & Kaminski, 2002). According to the instructional recommendations within the administration and scoring manual (Good, Simmons, Kame'enui et al., 2002, as cited in Good & Kaminski, 2002), students identifying 18 or more correct phonemes per minute are at "low risk" for future reading difficulties, while students identifying 7 or fewer correct phonemes per minute are "at risk" for future reading problems. According to the DIBELS Data System Technical

Report 1102 (Cummings, et al., 2011), the mean score for PSF during the kindergarten winter benchmark is 28.82 (17.78 SD).

According to the DIBELS technical manual (Dynamic Measurement Group, 2008), the reliability of the PSF ranges from .74 to .90. Concurrent and predictive validity are similarly strong. Concurrent measures range from .36 (TERA-3) to .59 (Nonsense Word Fluency task from DIBELS). Predictive validity measures range from .42 (Oral Reading Fluency task of DIBELS, as measured during the middle of second grade) to .55 (DRA Instructional Reading, as measured during the spring of first grade). Rouse and Fantuzzo (2006) demonstrated, by using a multivariate approach to demonstrate the unique contribution of both alphabet knowledge and phonological awareness to reading ability, that PSF was a strong predictor of first grade reading, vocabulary, and language.

DIBELS Instructional Recommendations. According to the instructional recommendations within the administration and scoring manual (Good, Simmons, Kame'enui et al., 2002, as cited in Good & Kaminski, 2002), individual student performance on DIBELS assessments can be analyzed on two levels: individual task performance and overall patterns of performance. On both the individual DIBELS tasks, as well as an overall pattern of performance, student outcomes are categorized into one of three outcome descriptors: low risk, some risk, or at risk. Alternative descriptors used to delineate outcome levels include deficit, emerging, and established. For purposes of this study, at risk and deficit are used interchangeably, as are emerging and some risk, and established and low risk.

These three ranges of performance were determined from analyses that utilized longitudinal predictive information from across year and within year comparisons, using data from all students within the DIBELS data system. Cutoffs were established primarily "based on the odds of achieving subsequent early literacy goals" and also with the percentage of students within in range being considered (Good, Simmons, Kame'enui et al., 2002). Rough targets were considered so that 20% of students in the broader sample were identified as at risk and another 20% of students were identified as some risk. Receiver Operator Characteristic (ROC) curves were created and examined for each individual measure and for the relevant subsequent benchmark goal. As an example, for the beginning of kindergarten ISF measure, four ROC curves were created for the four benchmark goals (ISF in the winter of kindergarten, PSF in the spring of kindergarten, NWF in the middle of first grade, and Oral Reading Fluency in the end of first grade). For each of these four measure and benchmark goals, two ROC curves were examined; one examining the odds in favor of achieving subsequent goals, and one examining the level of the goal with odds against achieving subsequent goals. These ROC curves, in addition to consideration of theoretical knowledge of beginning reading skill and reading skill growth, were used to establish a decision rule and instructional recommendation (Good, Simmons, Kame'enui et al., 2002).

When a student performs in the low risk range across several of the tasks assigned at any given benchmark period, the likelihood of being "on track" for meeting future reading achievement benchmarks is high, approximately 80%. Conversely, when a student's performance is in the deficit and/or at risk range, the odds are against achieving future goals; approximately 20% or fewer of students in the at risk range are expected to

achieve future reading goals. The third area of performance, identified as some risk or emerging, indicates that a clear prediction for future reading achievement was not possible, due to favors not being for or against, or 50 – 50 odds of obtaining future literacy goals. An example of this "some risk" identification at the beginning of kindergarten is a student who performs in the low risk range on ISF, but the at risk range on LNF (Good, Simmons, Kame'enui et al., 2002).

Wisconsin Knowledge and Concepts Examination (WKCE). The WKCE is a standardized, state mandated assessment of reading, language arts, science, social studies, and math. Reading portions of the WKCE are administered to students in the fall of third, fourth, fifth, sixth, seventh, eighth, and tenth grade in the state of Wisconsin. The WKCE, while developed specifically for the state of Wisconsin, is based from the *TerraNova* Comprehensive Test of Basic Skills, previously known as the California Achievement Tests or the CTB/McGraw-Hill (CTB/McGraw-Hill, 2003). The reading portion of the WKCE for grades three and four are "intact form[s] of the *TerraNova* [Third Edition]" (CTB/McGraw-Hill, 2003).

Items were developed to assess specific reading objectives: Word Meaning (approximately 25% of the third grade reading assessment), Understanding Text (approximately 30% of the third grade reading assessment, Analyze Text (approximately 35% of the third grade reading assessment), and Evaluating and Extending Text, (approxmiantely10% of the third grade reading assessment). Students are asked to read six to eight passages in which at least two passages are literary, two passages are informational text, one passage is poetry, and one passage is "everyday text" (such as charts, graphs, etc.). Students complete multiple choice and short answer (termed

constructed response) questions in one of the four basic areas (Wisconsin Department of Instruction (WI DPI) "Assessment framework for reading", 2006). According to the Fall 2009 WKCE Technical Manual (WI DPI & CTB/McGraw-Hill, 2010), the mean scaled score for third grade reading across the state of Wisconsin was 456.59 (39.43 SD).

Brown & Coughlin (2007) report that the reliability of the *TerraNova* demonstrates strong internal consistency, with coefficient values ranging from .80 to .95. The WKCE Technical Manual (WI DPI & CTB/McGraw-Hill, 2010), notes that internal consistency reliability estimates were generated using Cronbach's alpha coefficient and were reasonable for the number of items in the reading portion of the test. For the purposes of this study, the overall reading composite for third-grade was analyzed. Composite scaled scores can range from 270 to 640. Based on the criteria published (WI DPI "WKCE Scale Score ranges", n.d.), student scores were identified as "advanced" (scores of 507-640), as "proficient" (scores of 475-506), as "basic" (scores of 445-474), or as "minimal performance" (scores of 270-444).

Validity for the WKCE is noted to be 'adequate' (Anderson, 2010; Harwell, 2010). Construct validity was intensively measured by CTB-McGraw-Hill (WI DPI & CTB/McGraw-Hill, 2010). According to this report, in the area of reading, the correlations among content standards ranged from 0.56 to 0.84, which are interpreted as supporting the construct validity of the WKCE assessments. CTB/McGraw-Hill's linking study with the Pennsylvania System of School Assessments (as cited in Brown & Coughlin, 2007) indicated predictive relationships ranging from .67 to .82. Differential item functional analysis was completed for gender and ethnicity of respondents and results, and indicated "low to no [measurement bias] on items used in TerraNova"

(Anderson, 2010), which means the latent trait of students is not different as a function of gender, race, etc. Finally, significant care was taken to ensure items were not biased against any particular student groups (Anderson, 2010; WI DPI & CTB/McGraw-Hill, 2010).

Data Analysis

The data gathered was analyzed as follows for the following research questions. Statistical calculations were completed using SPSS (Version 18).

Question One. The first question sought to determine if a positive relationship existed between participation in a four-year-old kindergarten program (4K) and performance scores on early literacy measures gathered upon kindergarten entry, specifically the DIBELS tasks of Initial Sound Fluency (ISF) and Letter Naming Fluency (LNF). Two analyses compared 4K participants with non-participants; first, examining DIBELS measures of the 4K participants to a comparison group, called No 4K, and second, across the DIBELS measures of ISF and LNF within cohort. The comparison group analysis was completed using the winter benchmark data, due to a lack of data fall benchmark data from the school district. Specifically, tasks assessed at this time include Initial Sound Fluency (ISF), Letter Naming Fluency (LNF), Nonsense Word Fluency (NWF) and Phoneme Segmentation Fluency (PSF). Multivariate analyses of variance were completed for both the examination of within cohort differences as well as the comparison for the first 4K class to the No 4K comparison group.

Question Two. The second research question examined whether or not participation in 4K resulted in different reading risk status outcomes upon kindergarten entry. As noted, performance on DIBELS tasks are classified into three categories, at

risk, some risk, and low risk (Good, Simmons, Kame'enui et al., 2002). For the purposes of this study, scores were collapsed and used to create a dichotomous variable where at risk was coded as 1 and was equivalent to receiving an "at risk" benchmark outcome, and not at risk was coded as 0 and included outcomes of either 'low risk' or 'some risk' from the DIBELS benchmark assessment. Using the dichotomous "at risk" variable and 4K participation status, percentages of 4K participants and non participants being at risk or not at risk were calculated, using a chi-squared analysis.

Question Three. The third research question examined whether participants in 4K programming scored better than non-participants on the third grade WKCE reading assessment. Analyses compared 4K participants to non-participants within classes, as well as in comparison to a comparison group. Analyses of variance sought to determine whether participation or non-participation in the 4K program was associated with differences in performance on the WKCE reading assessments.

CHAPTER IV

Results

The relationship between literacy skills and participation in a school district supported universal four-year-old kindergarten (4K) program is presented in this chapter. Specifically, the relationship between participation in a 4K program and measures of early literacy upon entry in kindergarten and differences in performance between 4K participants and non-participants on state mandated assessments of reading in third grade were examined.

Data Screening

The data gathered for analysis included status of participation in the 4K program, scores from Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) measures completed in the fall and winter of Kindergarten, and Wisconsin Knowledge and Concepts Examination (WKCE; CTB/McGraw-Hill, 2005) reading scores from the beginning of third grade. Skewness and kurtosis statistics (Table 3) and sample distributions were reviewed to determine normality of the sample data. Both ISF Fall and NWF Winter demonstrated skewed distributions when a threshold of 1.0 was used to identify an unusual raw score distribution. Tabachnik and Fidell (2001) suggest that visual analysis of the shape of a sample distribution is recommended when skewness statistics are high for variables in a large sample. Visual analysis of box plots for ISF and NWF demonstrated that the high skewness scores were due to extremely high scores on each measure in the 4K participant sample and not due to significant outliers in the sample. Finally, assumptions were met for all multivariate tests in the current sample,

including multivariate normality, homogeneity of covariance, and independence of observations.

Table 3

Descriptive Statistics for DIBELS and WKCE Scores

Measure	M (SD)	Range	Skewness (SE)	Kurtosis (SE)
ISF Fall K	11.64 (9.2)	0-60	1.15 (.073)	2.19 (.146)
ISF Winter K	20.78 (11.9)	0-80	.932 (.069)	1.79 (.138)
LNF Fall K	19.10 (14.9)	0-76	.626 (.073)	-0.18 (.146)
LNF Winter K	35.59 (16.5)	0-101	.175 (.069)	0.17 (.138)
NWF Winter K	20.65 (19.1)	0-144	2.107 (.069)	7.90 (.138)
PSF Winter K	26.48 (16.7)	0-73	.101 (.069)	-0.71 (.138)
WKCE Reading Scaled Score	456.31 (42.3)	270-576	-1.4 (.099)	5.0 (.197)

No outliers were removed from the data set; however, several student populations were removed prior to data analysis. In the district of participation, children who were identified under special education prior to or during their 4K year were automatically enrolled in the 4K program, which caused an overrepresentation of special education students within the 4K participant group relative to non-participants. Therefore, their data were removed. Eighty-nine students were identified as receiving special education services during 4K. Similarly, students identified as English Language Learners (ELL) are also automatically enrolled in the 4K program, and thus their data were removed. Fifty-two students were identified as ELL during 4K. Data was reviewed for attrition between 4K and kindergarten and/or third grade. Specific numbers of students lost through special services and/or attrition is noted in the demographic Table 2, located within Chapter III.

Question One: Relationship Between 4K and Kindergarten Literacy Skills

Did the First 4K class score higher than the No 4K class? To examine the relationship between 4K participation and early literacy skills at kindergarten entry, comparisons were first made between a comparison group, a class of students who did not have 4K available prior to kindergarten (No 4K), and the first class of 4K participants (1^{st} 4K). This analysis is hypothesized to act as a quasi-control for a possible effect 4K availability may have had on all students, even those who did not directly participate. Due to district assessment schedules, DIBELS scores were not available for the beginning of Kindergarten for the No 4K cohort. Therefore, measures from the middle of Kindergarten (ISF, LNF, NWF, and PSF) were compared.

A one-way multivariate analysis of variance revealed a significant main effect for 4K participation. Wilks $\lambda = .924$, $F(4, 396) = 26.1$, $p < .001$, partial $\eta^2 = .076$. Please refer to Table 4 for means and standard deviations for each group. Given the significance of the overall test, the univariate main effects were examined. Significant univariate main effects for 4K participation were obtained for ISF, $F(1, 399) = 30.90$; $p < .001$; partial $\eta^2 = .072$, for LNF, $F(1, 399) = 6.43$; $p = .012$; partial $\eta^2 = .016$, and for PSF, $F(1, 399) = 6.74$; $p = .010$; partial $\eta^2 = .017$. Significant univariate main effects for 4K participation were not obtained for NWF, $F(1, 399) = 1.38$; $p = .24$; partial $\eta^2 = .003$. In order to make an alpha correction to account for multiple ANOVAs calculations, a Bonferoni correction was applied resulting in statistical significance being defined at $p < .01$.

Table 4

Mean and Standard Deviation of DIBELS for 4K Group (n = 148) and Comparison Group (n = 450)

DIBELS Measure	Group	M	SD
ISF	No 4K	15.93	9.66
	1st 4K	22.29	12.22
LNF	No 4K	33.49	15.88
	1st 4K	37.87	15.70
PSF	No 4K	23.72	15.86
	1st 4K	28.06	14.04
NWF	No 4K	19.17	16.85
	1st 4K	21.51	21.19

Did 4K participants score better than their non-participating classmates?

The means and standard deviations of DIBELS scores for 4K participants and nonparticipants are listed in Table 5. In order to determine whether differences in kindergarten DIBELS scores existed across kindergarten classes, data were analyzed to evaluate for a class effect. Significant differences between classes did not exist, nor was there an interaction between class and 4K participation. As no differences were found, the data was consolidated across all class years for the final analysis.

Table 5

Descriptive Statistics of DIBELS Scores for 4K Participants and Nonparticipants

Measure	Class	4K participation	Mean	SD
Initial Sound Fluency	1st 4K	yes (n=123)	12.62	9.667
		no (n=161)	10.70	9.996
	2nd 4K	yes (n=152)	13.50	9.667
		no (n=138)	10.8	8.325
	3rd 4K	yes (n=154)	13.84	9.904
		no (n=127)	9.12	7.613
	4th 4K	yes (n=183)	11.35	7.738
		no (n=82)	11.44	9.332
	Total	yes (n=612)	12.76	9.229
		no (n=508)	10.25	8.899
Letter Naming Fluency	1st 4K	yes (n=123)	21.59	14.159
		no (n=161)	14.34	13.997
	2nd 4K	yes (n=152)	23.78	14.600
		no (n=138)	16.92	15.945
	3rd 4K	yes (n=154)	22.85	15.225
		no (n=127)	15.23	14.826
	4th 4K	yes (n=183)	19.92	13.520
		no (n=82)	16.37	13.351
	Total	yes (n=612)	21.95	14.405
		no (n=508)	15.59	14.648

It was predicted that participation in the 4K program would be related to higher scores on an early literacy assessments conducted upon entry to kindergarten. A one-way MANOVA revealed a significant multivariate main effect for 4K participation. Wilks $\lambda = .955$, $F(2, 1111) = 26.1$, $p < .001$, partial $\eta^2 = .045$. Given the significant overall multivariate main effect, the univariate main effects were examined. Significant univariate main effects for 4K participation were obtained for ISF, $F(1, 1112) = 20.134$; $p < .001$; partial $\eta^2 = .018$; and LNF, $F(1, 1112) = 50.54$; $p < .001$; partial $\eta^2 = .043$. Children in the 4K program obtained higher scores on both the ISF and LNF tasks at kindergarten entry. In order to make an alpha correction to account for multiple

ANOVAs calculations, a Bonferoni correction was applied resulting in statistical significance being defined at $p < .025$. Assumptions were met for all multivariate tests in the current sample, including multivariate normality, homogeneity of covariance, and independence of observations.

Question Two: 4K Participation and Reading Risk Status

The second research question sought to examine whether a relationship existed between 4K participation and risk status for later reading difficulties. Frequency data in Table 6 shows that 21% of children who did not participate in 4K were "at risk" for future reading difficulties; whereas, only 4% of 4K participants were "at risk". Frequency data related to the number of 4K participants and non-participants found to be at-risk for future reading difficulties was used to evaluate the relationship between the two variables, 4K participation and reading risk status. The chi-square test for independence of groups was significant, χ^2 (1, 840) = 56.62; $p < .001$; $r^2 = .067$. A significantly higher percentage of 4K non-participants were at risk for reading difficulties upon kindergarten entry.

Table 6

Classification Table

	Low Risk	At Risk	Totals
No 4K	335	89	424
4K Participants	401	16	417
Totals	736	105	

Question Three: Relationship Between 4K Participation and Third Grade Reading Assessments

The final research question examined whether a difference existed between 4K participants and non-participants on measures of reading during third grade. WKCE reading scaled scores were gathered for 4K participants (n = 117) and non-participants (n = 309) for the first two classes of 4K participants. For 4K participants, the mean score was 459.5 (SD = 29.55). For non-participants, the mean score on the WKCE was 456.35 (47.06 SD). The ANOVA for differences in WKCE scaled scores was not significant, F (1, 424) = .455, p = .500 (η^2 = .001).

Chapter V

Discussion

Research indicates that students who demonstrate low early literacy skills in kindergarten and first grade will likely struggle to progress in reading skills at a similar rate to their peers (Burke et al., 2009) and therefore likely fall further behind academically (Good, Simmons, & Smith, 1998). State-funded or universal preschools have evolved as one of several approaches to improve academic outcomes for all students, particularly those identified with low socioeconomic or minority status. The present study sought to examine the relationship between participation in a universal preschool program and performance on measures of early literacy and reading, particularly for children who are more likely to be identified as 'at risk' for academic difficulties. This section will discuss the results and implications of the study. Limitations of the study and implications for future research will also be presented.

Relationship Between 4K Participation and Kindergarten Literacy Skills

Across the 4K evaluation literature, variability exists regarding the degree to which academic skills (language, reading, math, and print knowledge) are impacted by preschool participation. The first research question of this study asked whether children who participated in 4K demonstrated higher early literacy skills. Comparisons of the first class of 4K to a prior year (in which 4K was not available) were completed using the winter benchmark assessments measures from Dynamic Indicators of Early Literacy Skills (DIBELS), which included letter naming, phonological awareness, letter sound knowledge, and alphabetic principle. Analysis indicated that letter knowledge, letter sound knowledge, and alphabetic principle skills were stronger for the 4K participants.

There was not a significant difference between groups on other beginning reading skills (i.e., Nonsense Word Fluency).

Additional comparisons were made between 4K participants and non-participants across multiple cohorts of students and determined that 4K participants consistently performed higher than non-participants on measures of Initial Sound Fluency and Letter Naming Fluency in the fall of their kindergarten year. It should be noted that, although the results were noted to be significant in favor of 4K participants, the effect sizes were small for ISF and small to moderate for LNF. This indicates that the performance of the 4K participants was only slightly stronger than the non-participants, and that other confounding factors may account for this discrepancy, such as sample size.

Although no significant effects were identified for alphabetic principle (i.e., Nonsense Word Fluency), 4K participants in this sample did demonstrate stronger performance on other early reading skills (Initial Sound Fluency, Letter Naming Fluency, and Phoneme Segmentation Fluency). According to a meta-analysis of program evaluations (Gilliam & Zigler, 2001) this finding is consistent with studies in Georgia and Florida that demonstrated significant effects for 4K participants during kindergarten based on standardized achievement tests and developmental rating scales. Georgia and Florida were the only two states to specifically measure reading achievement during kindergarten. The current study adds to the limited investigations within the literature. Few programs, as reviewed in the meta-analysis (Gilliam & Zigler, 2001), specifically examined literacy skills at any grade level; of the two programs, only one found significant differences, and effect sizes were low.

The significant results from the current study also do not match results from the NIEER studies reviewed (Frede, Jung, Barnett, Lamy et al., 2007; Hustedt et al., 2008; Lamy, Barnett, & Jung, 2005a; Lamy, Barnett, & Jung, 2005b; Lamy, Barnett, & Jung, 2005c; Lamy, Barnett, & Jung, 2005d; Wong et al., 2007). In those studies, while significant effects were found in the areas of vocabulary, math, and print awareness, no significance was found in the area of phonological awareness. The method of assessment of these skills, using CBM benchmark measures, may account for the significant findings, as the NIEER studies used standardized assessment tasks to measure skills, and reported raw scores, which does not give an indication of whether all or any students were performing at an expected achievement or not.

4K Participation and Reading Risk Status

Individual DIBELS subtest scores from each benchmark have the ability to be combined and compared to data from the DIBELS authors to create a prediction about a student's future reading performance. This prediction, labeled as instructional recommendation, can be used to plan for interventions or instruction. This study sought to determine if there was a relationship between 4K participation and reading risk status, specifically whether participant in 4K resulted in lower risk for reading difficulties. Results of the statistical analysis indicate a moderate effect size; children who did not participate in the 4K program were more likely to be rated as "at risk" for reading difficulties at kindergarten entry. This research question adds a new way of examining the effects of preschool participation to the literature. CBMs, by design, are predictive of future academic performance. By using CBMs to measure student likelihood of future academic difficulties, researchers are able to examine the effectiveness of preschool

programming in a new way that can potentially provide earlier indicators of student risk, which can in turn provide earlier intervention and support.

Relationship between 4K and Third Grade Reading Assessments

The third research question sought to examine the relationship between 4K participation and performance on state-wide assessments of reading completed in third grade. Specifically, an examination of whether or not there was a difference between 4K participants and non-participants on the overall scaled score of the reading assessment portion of the WKCE was examined. Analyses indicate that there was not a significant difference between 4K participants and non-participants on third grade assessments of reading achievement. This was true for five other program evaluations, with significant effects found in only two states (Gilliam & Zigler, 2001). These results suggest that, as children progress through school, the differences between participants and nonparticipants are less evident.

While some researchers may argue that this 'fading' of effects may discourage the practice of early childhood education programs, these results could also be interpreted as having "evened the playing field" for preschool participants. The 4K program experience may have provided participants with opportunities to build the early literacy skills that are important for early reading development. The same students may have scored lower on third grade measures had they not attended the 4K program.

The current study adds to the research about the effectiveness of preschool programming on the achievement of participants upon entry to kindergarten. However, the results also demonstrate that preschool programming or early intervention in general, is only a part of the greater solution for addressing the achievement gap. Preschool

programming is not a magic cure for this educational and societal issue. Rather, high-quality preschool must be accompanied by high-quality elementary school practices in order to truly address this concern and improve the education outcomes of all children (Anderson, Shinn, Fullilove, Scrimshaw, Fielding, Normand, et al., 2003; Magnuson & Waldfogel, 2005; Ramey & Ramey, 2004).

Implications

Overall, the results of the current study suggest that 4K participants did demonstrate stronger early literacy skills than non-participants upon kindergarten entry. Studies show that students exhibiting deficits in pre-reading skills upon kindergarten entry often struggle to progress in reading skills at a similar rate to their peers as they move through the elementary years (Burke et al., 2009). Even students receiving additional educational supports to remediate their reading skills struggle to 'catch up', due to a combination of a low level of initial skills and slow rate of skill growth (Good, Simmons, & Smith, 1998). Research suggests that early intervention, or prevention, is more effective than interventions occurring at a later grade level (Burke et al., 2009; Cunningham & Stanovich, 1997; Juel, 1988; Scarbourough & Parker, 2003; Stanovich, 1986), and preschool programs are able to provide this early identification and early intervention to work toward stronger student performance (Anderson, Shinn, Fullilove, Scrimshaw, Fielding, Normand, et al., 2003; Magnuson & Waldfogel, 2005; Ramey & Ramey, 2004).

This study evaluated the relationship between 4K participation and academic performance with a new approach; specifically, the use of CBMs as indicators of risk status. In the studies reviewed (Frede, Jung, Barnett, Lamy et al., 2007; Hustedt et al.,

2008; Lamy, Barnett, & Jung, 2005a; Lamy, Barnett, & Jung, 2005b; Lamy, Barnett, & Jung, 2005c; Lamy, Barnett, & Jung, 2005d; Wong et al., 2007), standardized assessments were used to measure the skills of the study participants, and the raw scores from each assessment from each child were used to determine effect sizes. The inherent problem with this method of analysis is that the actual proficiency level of the study participants is unknown. In using the standardized raw scores style of data comparison, it is possible to see significant differences and effect sizes between state-funded preschool participants and nonparticipants, while both groups may have still performed below the desired level of achievement for their age. By using the CBM measures, researchers were able to compare the differences between 4K participants and nonparticipants, and also compare the scores of either group to a criterion of performance.

This adds information to the method with which schools can evaluate 4K effectiveness at a local level. CBM measures are commonly being used by local schools to identify students in need of early intervention, and because many of these schools are also offering 4K programming, methods of program evaluation, such as those used in the current study, can be easily implemented by districts and evaluators. The concept of 'risk estimates', as thoroughly described in the DIBELS Administration Manual (Good & Kaminski, 2002), is also an important consideration in program evaluation of 4K programming. Using instructional recommendation methods will allow for educators, even at entry to kindergarten, to identify children who are at risk for academic failure and provide interventions. Information concerning students entering kindergarten can be used to offer ideas for considering where and how to most effectively intervene and collaborate with community partners and programs to support children and families prior

to entering school. In other words, this CBM data can be used to identify the programs, neighborhoods, and schools that are in greater need for supports, services, and additional opportunities.

The current study allowed for a more precise analysis of 4K on students with low socioeconomic status. Although the 4K program in the district collaborating in this study is a universal program, and thus available to all four year old children within the district, the elementary schools with the highest levels of free and reduced lunch eligibility (and thus low socioeconomic status) in the district were used for data collection and comparison. This allowed researchers to compare participants and nonparticipants with less concern of low socioeconomic status making a significant difference on effects. In addition, the targeted analysis of the students with the highest rates of free and reduced lunch (or low socioeconomic status) allowed the district to determine if the 4K program effected the neediest populations in the district. This practice can be repeated in future studies, either comparing students identified with a low socioeconomic status to each other, or comparing students not identified with a low socioeconomic status to each other.

Finally, the current study has specific relevancy to the field of school psychology. First, school psychologists should advocate for high quality early childhood opportunities. The current study provides additional support for the practice of early intervention. Second, National Association of School Psychologists' Strategic Plan goals of enhancing professional competency and advocacy for all children (National Association of School Psychologists, 2008) suggest that school psychologists should be familiar with ways that their training can support advocacy efforts for all children. This study highlights the relative ease with which preschool evaluations can be completed

with the use of CBM and school district data. The use of evidence-based practices towards evaluation and accountability for children of all ages is also directly aligned with the National Association of School Psychologist' policies (National Association of School Psychologists, 2008).

Limitations and Future Research

Several limitations are present in the current study. Sample selection presents one limitation area. First, student data were gathered after children had already been enrolled and participated (or did not participate) in the 4K program, thus the groups of students were naturally occurring, and were not controlled for. The studies reviewed that used the Regression Discontinuity Design (RDD), sought to address this selection bias by comparing children who had completed the preschool program and entering kindergarten to those children who had not yet received the 'treatment' of the preschool program by adhering to strict age guidelines for preschool and kindergarten enrollment (Frede, Jung, Barnett, Lamy et al., 2007; Hustedt et al., 2008; Lamy, Barnett, & Jung, 2005a; Lamy, Barnett, & Jung, 2005b; Lamy, Barnett, & Jung, 2005c; Lamy, Barnett, & Jung, 2005d; Wong et al., 2007). However, this method requires both groups of children to complete the same measures at the same point in time. In the current study, the No 4K cohort, was used as a type of comparison group. Due to limitations with demographic information, the RDD design could not be incorporated into the current study.

Second, the "no 4K" comparison groups were not a matched control group and students were not randomly assigned to 4K and no 4K groups. Because of this, it is impossible to interpret a causal link between the 4K program (the treatment) and the outcomes of stronger skills on particular literacy measures. Multiple other factors,

including gender, cultural factors, specific age, prior life experiences, level of family involvement, and specific family socioeconomic status (among other variables) may confound these results. Using a matched control group for comparison would have lent more strength to comparisons between groups. While significant differences were determined to exist between 4K participant and nonparticipants, there was no data collected on the curriculum and instructional practices used in each individual preschool. Quality of 4K programming has been identified as an essential component of 4K program evaluations and important for consideration when reviewing program effectiveness (Espinosa, 2002; Frede, 1995). Because there was no data collected to determine the quality of programming, it is impossible to know what type of impact this may or may not have had on the performance of student skills.

A final issue with sample selection was missing demographic information. Free and reduced lunch status, which is commonly used by schools and researchers as an indicator of low socioeconomic status, was not available at the subject level, and rather was reported only by percentage of the entire elementary school. Other demographic information missing was the types of alternate experiences afforded to those students who did not participate in 4K. In other words, the sample technique of the current study does not allow us to know what experiences or preschool programmings were afforded to the 4K non-participants. An informal survey conducted by the district after the first 4K year determined that, of the 141 parents responding to the reason(s) that they did not enroll their child in 4K in the first year of the program, 68 percent indicated that they had attended another preschool site, 24 percent were listed as at home with the parent, 17 percent were noted as in an 'in-home day care' setting, and 8 percent were in a day care

setting. This data shows that many nonparticipants of the 4K program were enrolled in some other preschool-type program.

Another limitation to the study is the lack of a direct link of risk status as identified by DIBELS and the accuracy of this prediction to the WKCE. For purposes of this study, general cut scores, as recommended within the DIBELS Administration and Scoring manual (Good & Kaminski, 2002), were used to determine risk status. However, these cut scores were not directly assessed for accurately predicting risk on the WKCE specifically.

As preschool programming continues to develop, and the evaluations of the effectiveness of these programs continue to evolve, there are many areas for future research. Due to the significant variability across program models, it will be important to measure effectiveness of specific data; for example, curriculum used, hours of instruction, class size, and student to teacher ratios and their possible impacts on school achievement are all areas in which further research should be conducted. In addition, educational success is more than just proficient academic skills. Social emotional functioning, behavioral skills, student health, grade retention, special education eligibility, independence/self-help skills, parent involvement, and other areas are all important factors that should be considered when determining whether or not a preschool program is 'effective' in addressing student readiness.

The skills being used to measure future achievement, and the way they are measured, should also be further researched. For example, the NICHD National Reading Panel report (2001) documented that phonemic awareness is one of the important skills needed for future reading success. However, many of the studies reviewed did not find

significant effect sizes on measures of phonological awareness when comparing preschool participants to non-participants. Further research is needed to determine whether phonemic awareness skills for participants are truly not affected by preschool experiences, or, if the ways in which these skills are being measured are not capturing the difference in knowledge and application ability for participants versus nonparticipants.

Further research is also important for determining the various methods needed to close the achievement gap. Results of studies suggest that preschool is an effective intervention in that participants demonstrate some areas of stronger skills than non-participants, but that commonly the positive effects are noted to fade after several years. Additional research can help to determine the possible causes for this fading, or can work to determine the educational practices needed in the primary grades to support preschool practices to avoid the lessening of preschool effects.

Finally, additional research is needed to support the idea of stronger effects from more preschool experience. Frede, Jung, Barnett, & Figueras (2009) demonstrated that children who attended two years of the program demonstrated stronger initial effect sizes across multiple skill areas and that these skills continued to be stronger than nonparticipants for a longer period of time. Further research to support or negate this claim will be important as policy and funding for preschool programming continues to change across the nation.

Conclusion

Preschool programming evaluation and effectiveness continues to be an important area of research in the field of education and school readiness. Early intervention continues to be a strong factor in bringing about student success, particularly for children

at risk for academic failure. This early intervention needs to be of high-quality, and needs to be systematically monitored for effectiveness, and modified or changed to meet the needs of the community. The methods of evaluating preschool programming needs to continue to be challenged, in order to assure the programming is bringing about desired effects in the most efficient and economical ways possible. By consistently evaluating programming, we are providing the best ways to ensure student educational success in preschool and beyond.

References

Adams, M.J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press

Anderson, J.O. (2010). [Review of the TerraNova, Third ed.]. In *The eighteenth mental measurements yearbook.* Available from http://www.unl.edu/buros/

Anderson, L.M., Shinn, C., Fullilove, M.T., Scrimshaw, S.C., Fielding, J.E., Normand, J., ... Task Force on Community Preventative Services. (2003). The effectiveness of early childhood developmental programs: A systematic review. *American Journal of Preventative Medicine, 24* (3S), 32-46.

Barnett, W.S., Brown, K., & Shore, R. (2004). *The universal versus targeted debate: Should the United States have preschool for all?* New Brunswick, NJ: National Institute for Early Education Research.

Barnett W.S., Carolan, M.E., Fitzgerald, J., & Squires, J.H. (2012). *The state of preschool 2012: State preschool yearbook.* New Brunswick, NJ: National Institute of Early Education Research.

Barnett, W.S., Hustedt, J.T., Robin, K.B., & Schulman, K.L. (2005). *The state of preschool: 2005 state preschool yearbook.* New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W.S., Lamy, C., & Jung, K. (2005). *The effects of state prekindergarten programs on young children's school readiness in five states.* New Brunswick, NJ: National Institute of Early Education Research.

Barnett, W.S., Robin, K.B., Hustedt, J.T., & Schulman, K.L. (2003). *The state of preschool: 2003 yearbook.* New Brunswick, NJ: National Institute of Early Education Research.

Brown, R.S. & Coughlin, E. (2007). *The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region* (Issues & Answers Report, REL 2007-No.017). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from http://ies.ed.gov/ncee/edlabs

Burke, M.D., Hagan-Burke, S., Kwok, O., & Parker, (2009). Predictive validity of early literacy indicators from the middle of kindergarten to second grade. *Journal of Special Education, 42* (4), 209-226. doi: 10.1177/0022466907313347

Burke, M.D., Hagan-Burke, S., Zou, Y., & Kwok, O. (2010). A structural equation model using fluency-based early literacy measures to predict emerging reading ability in kindergarten. *Remedial and Special Education, 3*(5), 385-399. doi: 10.177/0741932509355949

Campbell, F. A. & Ramey, C.T. (1994). Effects of early intervention on intellectual and achievement: A follow up study of children from low-income families. *Child Development, 65*, 684-698.

Campbell, F.A., Ramey, C.T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6*(1), 42-57.

Carver, R.P. (1974). Two dimensions of tests: Psychometric and edumetric. *American Psychologist, 29*(7), 512-518. doi: 10.1037/h0036782

Chall, J.S. (1967). *Learning to read: The great debate*. New York, NY: McGraw-Hill.

CTB/McGraw-Hill (2003). *An investigation of score comparability on the Wisconsin Knowledge and Concepts Examinations (WKCE) and TerraNova Tests* (CTB/McGraw-Hill Research Report). Available at http://oea.dpi.wi.gov/files/oea/doc/ctbalign03.doc

Cummings, K.D., Otterstedt, J., Kennedy, P.C., Baker, S.K., & Kame'enui, E. J. (2011). *DIBELS Data System: 2009-2010 percentile ranks for DIBELS 6th Edition benchmark assessments* (Technical Report 1102). Eugene, OR: University of Oregon Center on Teaching and Learning.

Cunningham, A.E., & Stanovich, K.E. (1997). Early reading acquisition and its relations to reading experience and ability ten years later. *Developmental Psychology, 33*(6), 934-945.

Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives, 15*(2), 213-238.

Dunn, L.M. & Dunn, L.M. (1997). *Peabody picture vocabulary test*, 3rd edition. Circle Pines, MN: American Guidance Service.

Dynamic Measurement Group (2008). *DIBELS 6th Edition technical adequacy information* (Tech. Rep. No. 6). Eugene, OR: Author. Available at http://dibels.org/pubs.html

Early, D., Barbarin, O., Barnett, W.S., Bryant, D., Burchinal, M., Chang, F., … Weaver, W. (2005). *Pre-kindergarten in eleven states: NCELD's Multi-State Study of Pre-*

Kindergarten & Study of State-Wide Early Education Programs (SWEEP). Completed by the National Center for Early Development and Learning organization. Retrieved from University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Institute website: http://www.fpg.unc.edu/sites/default/files/resources/reports-and-policy-briefs/NCELD_PreK-in-eleven-states_Working-Paper_2005.pdf

Espinosa, L.M. (2002). *High-quality preschool: Why we need it and what it looks like.* Available at http://nieer.org/resources/policybriefs/1.pdf

Frede, E.C. (1995). The role of program quality in producing early childhood program benefits. *Future of Children, 5*(3), 115-132.

Frede, E., Jung, K., Barnett, W.S., & Figueras, A. (2009). *The APPLES blossom: Abbott Preschool Program Longitudinal Effects Study (APPLES) preliminary results through second grade.* New Brunswick, NJ: National Institute of Early Education Research.

Frede, E., Jung, K., Barnett, W.S., Lamy, C.E., & Figueras, A. (2007). *The Abbott Preschool Program Longitudinal Effects Study (APPLES).* New Brunswick, NJ: National Institute of Early Education Research.

Fuchs, D. & Fuchs, L.S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93-99. doi:10.1598/RRQ.41.1.4

Gilliam, W.S. & Zigler, E.F. (2001). A critical meta-analysis of all impact evaluations of state-funded preschools from 1977 to 1998: Implications for policy, service delivery, and program evaluation. *Early Childhood Research Quarterly, 15*, 441-473.

Goffreda, C.T. & DiPerna, J.C. (2010). An empirical review of psychometric evidence for the Dynamic Indicators of Basic Early Literacy Skills. *School Psychology Review, 39*(3), 463-483.

Good, R.H., Gruba, J., & Kaminski, R.A. (2001). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.) *Best Practices in School Psychology IV* (pp.679-700). Washington, DC: National Association of School Psychologists.

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.uorgeon.edu/

Good, R.H., Simmons, D., Kame'enui, E., Kaminski, R.A., & Wallin, J. (2002). DIBELS instructional recommendations: Intensive, strategic, & benchmark. In R.H. Good & R.A. Kaminski (Eds.), *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available: http://dibels.uoregon.edu/

Good, R.H., Simmons, D.C., & Smith, S.B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review, 27*(1), 45-56.

Gormley, W.T. & Gayer, T. (2003). *Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-k program.* (CROCUS Working Paper No. 1). Washington, DC: Georgetown University, Center for Research on Children in the U.S.

Grossen, B. (1997). *A synthesis of research on reading from the National Institute of Child Health and Human Development.* Retrieved from http://www.nrrf.org/synthesis_research.htm

Harwell, M. (2010). [Review of the TerraNova, Third Edition]. In *The eighteenth mental measurements yearbook.* Available from http://www.unl.edu/buros/

Howse, R.B., Lange, G., Farran, D.C., & Boyles, C.D. (2003). Motivation and self-regulation as predictors of achievement in economically disadvantaged young children. *Journal of Experimental Education, 71*(2), 151-174.

Huffman, L.C., Mehlinger, S.L., & Kerivan, A.S. (2000). *Risk factors for academic and behavioral problems at the beginning of school.* Bethesda, MD: National Institute of Mental Health.

Hustedt, J.T., Barnett, W.S., Jung, K., & Figueras, A. (2008). Impacts of New Mexico prek on children's school readiness at kindergarten entry: Results from the second year of a growing initiative. Retrieved from http://nieer.org/resources/research/NewMexicoRDD0608.pdf

Imbens, G.W. & Lemieux, T. (2008). Regression discontinuity design: A guide to practice. *Journal of Econometrics, 142*, 615-635.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437-447.

Kaminski, R., Cummings, K.D., Powell-Smith, K.A., & Good, R.H. (2008). Best practices in using Dynamic Indicators of Basic Early Literacy Skills for formative assessment and evaluation. *Best Practices in School Psychology V, 4*, 1181-1204.

Kame'enui, E. J. (Ed.). (1996). *Learning to read/reading to learn: Helping children with learning disabilities to succeed information kit* (Technical Report). Eugene, OR: National Center to Improve the Tools of Educators.

Kirby, J.R., Parrila, R.K., & Pfiffer, S.L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95*(3), 453-464. doi: 10.1037/0022-0663.95.3.453

Lamy, C., Barnett, W.S., & Jung, K. (2005a). *The effects of the Michigan School Readiness Program on young children's abilities at kindergarten entry*. Retrieved from http://www.highscope.org/file/research/mi.pdf

Lamy, C., Barnett, W.S., & Jung, K. (2005b). *The effects of South Carolina's early childhood programs on young children's readiness*. New Brunswick, NJ: National Institute of Early Education Research.

Lamy, C., Barnett, W.S., & Jung, K. (2005c). *The effects of West Virginia's early education program on young children's school readiness*. New Brunswick, NJ: National Institute of Early Education Research.

Lamy, C., Barnett, W.S., & Jung, K. (2005d). *The effects of Oklahoma's Early Childhood Four-Year-Old Program on young children's school readiness*. New Brunswick, NJ: National Institute of Early Education Research.

Lonigan, C.J., Wagner, R.K., Torgesen, J.K., & Rashotte, C.A. (2002). *Preschool Comprehensive Test of Phonological and Print Processing*. Austin, TX: ProEd.

Lonigan, C.J., Wagner, R.K., Torgesen, J.K., & Rashotte, C.A. (2007). *TOPEL: Test of Preschool Early Literacy*. Austin, TX: PRO-ED.

Magnuson, K.A., Meyers, M.K., Ruhm, C.J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research, 41*(1), 115-157. doi: 10.3102/00028312041001115

Magnuson, K.A. & Waldfogel, J. (2005). Early childhood care and education: Effects on ethnic and racial gaps in school readiness. *Future of Children, 15*(1), 169-196. doi: 10.1353/foc.2005.0005

Masse, L.N. & Barnett, W.S. (2002). *A benefit cost analysis of the Abecedarian Early Childhood Intervention.* New Brunswick, NJ: National Institute of early Education Research. Available at http://www.nieer.org

National Association of School Psychologists. (2008). *Early childhood services* (Position Statement). Bethesda, MD: Author.

National Institute of Child Health and Development (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: US Government Printing Office.

National Institute of Child Health and Development Early Childcare Research Network (2003). Does quality of child care affect child outcomes at age 4 ½? *Developmental Psychology, 39*, 451-469.

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 et Seq. (West 2003).

Puma, M., Bell, S., Cook, R., Heidi, C., Broene, P., Jenkins, F., … Downer, J. (2010). *Third grade follow up to the Head Start Impact Study Final Report* (OPRE Report No. 2012-45). Washington, DC: Office of Planning, Research, & Evaluation,

Administration for Children and Families, U.S. Department of Health and Human Services.

Ramey, C.T. & Ramey, S.L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merrill-Palmer Quarterly, 50*(4), 471-491. doi: 10.1353/mpq.2004.0034

Reynolds, A.J. (2000). *Success in early intervention: The Chicago Child-Parent Centers.* Lincoln, NE: University of Nebraska Press.

Reynolds, A.J., Temple, J.A., Ou, S., Robertson, D.L., Mersky, J., Topitzes, J.W., & Niles, M.D, (2007, March). *Effects of a preschool and school-age intervention on adult health and well-being: Evidence from the Chicago longitudinal study.* Paper presented at the biennial meeting of the Society for Research in Child Development, Boston, MA. Retrieved from http://www.plan4preschool.org/documents/pk-intervention. pdf

Rouse, H.L. & Fantuzzo, J.W. (2006). Validity of the Dynamic Indicators for Basic Early Literacy Skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review, 35*(3), 341-355.

Scarborough, H.S. & Parker, J.D. (2003). Matthew effects in children with learning disabilities: Development of reading, IQ, and psycho-social problems from grade 2 to grade 8. *Annals of Dyslexia, 53*(1), 47-71.

Schweinhart, L.W., Barnes, H.V., & Weikart, D.P. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27.* Retrieved from http://www.aypf. org/publications/nomoreisle/PDF/108-110.pdf

Schweinhart, L.J., Montie, J., Xiang, Z., Barnett, W.S., Belfield, C.R., & Nores, M. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through age 40* (Monographs of the High/Scope Educational Research Foundation, 14). Ypsilanti, MI: High/Scope Educational Research Foundation.

Shinn, M.R., Good, R.H., Knutson, N., Tilly, W.D., & Collins, V. (1992). Curriculum-based reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*(3), 458-478.

Speece, D.L., Mills, C., Ritchey, K.D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *Journal of Special Education, 36*(4), 223-233. doi: 10.1177/00224669030360040

Snow, C.E., Burns, M.S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360-407.

Sugai, G. & Horner, R. (2002). The evolution of discipline practices: School-Wide Positive Behavior Supports. *Child and Family Behavior Therapy, 24*(1/2), 23-50.

Tabachnik, B.G. & Fidell, L.S. (2001). *Using multivariate statistics, Fourth ed.* New York: Allyn & Bacon.

Temple, J.A. & Reynolds, A.J. (2007). Benefits and costs of investments in preschool education: Evidence from the Child-Parent Centers and related programs. *Economics of Education Review, 26*(1), 126-144.

Thistlethwaite, D.L. & Campbell, D.T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology, 51*(6), 309-317.

Torgeson, J.K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*, 7-26.

U.S. Department of Education, Institute of Educational Sciences, National Assessment Governing Board, National Center for Education Statistics, National Association of Education Progress (2005). *The nation's report card.* Retrieved from http://nationsreportcard.gov/

U.S. Department of Education, Institute of Educational Sciences, National Assessment Governing Board, National Center for Education Statistics, National Association of Education Progress (2011). *The nation's report card.* Retrieved from http://nationsreportcard.gov/

U.S. Department of Health and Human Services, Administration of Children and Families (May 2005). *Head Start impact study: First year findings*. Washington, DC.

Walker, H.M. & Shinn, M.R. (2002). Interventions for achieving prevention goals and outcomes for at-risk children and youth. In M.R. Shinn, H.M. Walker, & G. Stoner (Eds.) *Interventions for academic and behavior problems: Preventative and remedial approaches* (pp 1-26). Bethesda, MD: National Association of School Psychology.

Wayman, M.M., Wallace, T., Wiley, H.I., Ticha, R., & Espin, C.A. (2007). Literature synthesis on curriculum-based measurements in reading. *Journal of Special Education, 41*(2), 85-120. doi: 10.1177/00224669070410020401

Wisconsin Department of Instruction. (2006). *Assessment Framework for Reading.* Available at http://oea.dpi.wi.gov/files/oea/pdf/read_framework.pdf

Wisconsin Department of Instruction. (n.d.). *WKCE Scale Score Ranges.* Available at http://oea.dpi.wi.gov/oea_wkce-crtcuts

Wisconsin Department of Instruction & CTB/McGraw-Hill. (2010). *Fall 2009 WKCE Technical Manual.* Available at http://oea.dpi.wi.gov/files/oea/pdf/td-2009-techman.pdf

Wong, V.C., Cook, T.D., Barnett, W.S., & Jung, K. (2007). *An effectiveness-based evaluation of five state prekindergarten programs using regression-discontinuity.* Retrieved from http://nieer.org/resources/research/EvaluationFiveStates.pdf

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

Zigler, E., Gilliam, W.S., & Jones, S.M. (2006). *A vision for universal preschool education.* Cambridge University Press.