Summer 2013

The Effect of Public Pre-Kindergarten Attendance on First Grade Reading and Social Achievement: A District Level Analysis

Michael James Haslip
Old Dominion University

THE EFFECT OF PUBLIC PRE-KINDERGARTEN ATTENDANCE ON FIRST
GRADE READING AND SOCIAL ACHIEVEMENT:
A DISTRICT LEVEL ANALYSIS

by

Michael James Haslip
B.A. October 2005, Dalhousie University, Canada
M.S. May 2008, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of
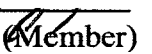
DOCTOR OF PHILOSOPHY

EARLY CHILDHOOD EDUCATION

OLD DOMINION UNIVERSITY
August 2013

Approved by:

Dr. Katharine Kersey (Chair)

Dr. Angela Eckhoff (Member)

Dr. Steve Myran (Member)

ABSTRACT

THE EFFECT OF PUBLIC PRE-KINDERGARTEN ATTENDANCE ON FIRST GRADE READING AND SOCIAL ACHIEVEMENT: A DISTRICT LEVEL ANALYSIS

Michael James Haslip
Old Dominion University, 2013
Chair: Dr. Katherine C. Kersey

The purpose of this study was to test the relationship between attending a school district Pre-K program and children's later literacy and behavior outcomes in the beginning and middle of first grade. Children's text level, letter-sound identification, sight words, spelling and citizenship skills were measured. The study included 880 children who attended the division Pre-K program and 176 children who attended no formal or institutional preschool of any kind. Selection bias was addressed through propensity score matching, completed using optimal matching. Independent-samples t tests were run on literacy and behavior measures. Subgroup analysis was performed to test the effect of Pre-K attendance on first grade behavior across different levels of school quality experienced in kindergarten.

Literacy measures in first grade included the *Phonological Awareness Literacy Screening* (PALS) and the *Developmental Reading Assessment, 2nd Edition* (DRA 2). Behavior was measured using a single behavior sum score composed of seven citizenship grades drawn from report cards.

The study found significant effects of attending Pre-K, in the small to medium range, on all reading measures administered in both the beginning and the middle of first

grade. The average effect size across all literacy measures was .35, a small to moderate effect. Behavior results for the two groups were nearly identical. Further subgroup analysis by school quality found that children who attended New Day Pre-K and also went on to a "fair quality" school in kindergarten, as opposed to a low quality school, had significantly better behavior than children who had no preschool experience.

The study strengthens the evidence supporting the ability of Pre-K to significantly improve children's literacy results with persistent gains into the middle of first grade. The study also raises the concern that children's social and emotional development is not receiving adequate instructional focus in the form of lesson plans, activities or time in the division Pre-K in order to make a meaningful difference for children in later grade school. The study contributes to the PK – 3 alignment discourse and offers a valuable case study in universal Pre-K access. Implications are shared for practice, policy and research.

Dedicated to the perseverance that is needed to close one's eyes to the surface world of events and behaviors, and instead to look within the sanctuary of the heart for the inspiration and beauty that transforms our world.

Bring forth an inspired world.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

ix

LIST OF TABLES

Table   Page

CHAPTER 1

INTRODUCTION TO THE STUDY

Among publicly funded preschool options, public pre-kindergarten (Pre-K) serves more children in the United States than any other type of preschool program, including Head Start (Barnett, Carolan, Fitzgerald, & Squires, 2011). Pre-K was the fastest growing preschool movement in the United States for ten years reaching an enrollment of 1.4 million children by 2007 (Barnett et al., 2008), until enrollment growth stopped as a result of the financial crisis in 2008. Currently, 39 states provide Pre-K programs. As states have added Pre-K, stakeholders such as state departments of education, early childhood research institutes and policy-makers, as well as local districts and early childhood researchers have initiated evaluative studies on their effectiveness at raising child school readiness scores and social development.

The literature on the longitudinal effect of Pre-K seeks to determine realistic expectations for the influence of Pre-K attendance, to investigate the perceived quality of such Pre-K programs, to highlight challenges in Pre-K and K – 12 alignment, and to evaluate later school quality in a framework of PK - 3 education where the intent is to sustain preschool gains for disadvantaged children into elementary school (Reynolds, Magnuson, & Ou, 2010). However, due to limited research, the potential ability of Pre-K attendance to make a significant difference on children's later academic and social achievement into the K – 3 years is not well understood.

Hindering these efforts are key challenges that face longitudinal Pre-K researchers relating to differing data systems, different assessment measures, and varying state

standards, among other issues (Hernandez, 2012). Methodological challenges also abound. For example, most of the Pre-K longitudinal studies have not identified the curricula being used by the Pre-K program, have not controlled for later school quality in the designs, and have given little attention to affective outcome measures. Adequate group equivalence through strong matched-pair designs remains a concern (see chapter 2 for an overview of the literature). Therefore, conducting a Pre-K effect study with significant controls, a matched-pair design with a large sample, and later school quality control, would improve the quality of evidence about the ability of Pre-K to sustain later gains and provide greater insight into the conditions that may make this possible.

Statement of the Problem

The disparity between low income and middle class children has negative consequences (Knudsen, Heckman, Cameron, & Shonkoff, 2006) including an achievement gap at school entry, reduced cognitive development, a lower achievement trajectory in later grades and increased delinquency and crime in later years. Vulnerable children are often resilient and tenacious in the face of adversity (Luthar, 2003) but opportunities to develop to their full potential are limited. Publicly funded preschool, as an intensive intervention for children from poverty, has been extensively researched over several decades and can be an effective support for child development (Camilli, Vargas, Ryan, & Barnett, 2010). Education and training in the earliest years significantly impacts' well-being in later life, in higher education, in adult health and income (Knudsen et al., 2006).

Public preschool, like Head Start, targeted Pre-K, or comprehensive PK-3 programs, serves the child and the society in a variety of ways, such as increased productivity (Heckman & Masterov, 2007) . Public involvement in preschool recognizes that intervention does help close the achievement gap, support families to improve child health, facilitate emotional development, reduce criminality and strengthen achievement (Temple, Arteaga, & Reynolds, 2010).

Research suggests that only high quality interventions are capable of making up these differences for vulnerable children (Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2011). Sustained quality must be experienced year over year, calling for partnership among agencies and alignment of the PK – 3 continuum (Reynolds, Magnuson, & Ou, 2010). Towards these ends, some researchers have sought to understand the relationship between Pre-K attendance and later school outcomes, beyond kindergarten entry (Magnuson, Ruhm, & Waldfogel, 2007) to address the question of Pre-K quality, PK – 3 alignment and later school quality. A common goal is to help close the achievement gap by ensuring that vulnerable children, and all children, receive a high quality early education from preschool through third grade and beyond.

Purpose of the Study

To contribute to an empirical understanding of the possible relationship between Pre-K attendance and later child outcomes, this study chose as its research setting a large state-funded pre-kindergarten program operated in an urban school district in the southeastern United States. The purpose of this study was to examine how attendance in the public school district's Pre-K program effected children's later academic and social

achievement in first grade. The study sought to contribute to our understanding of the sustainability of Pre-K impacts, maintain a balance between cognitive and affective measures, and examine the issue of later school quality in relation to Pre-K attendance, which has received almost no attention in the current literature, as the literature review shows.

District Demographics and Characteristics

This school division serves nearly 30,000 students from Pre-K through high school. The division operates twenty four elementary schools, fourteen of which are Title I. The division operates four unique school sites that are used exclusively for the preschool program. The four early childhood centers combine to serve close to 2000 preschoolers per year.

Fifty-four percent of students are Black, 28% are White, 11% are Hispanic and seven percent are mixed or other races. For current first grade students, these numbers are slightly different: Black (51%), White (25%), Mixed (20%), Asian (3%). About 58% of students (PK – 12) are economically disadvantaged. Poverty rates among young children are typically higher than that of older children across the United States, which is reflected in this city. Sixty-seven percent of first grade children in the city (2012 – 2013) received free (60%) or reduced (7%) lunches, meaning that just a third of all children pay full price for lunch. Higher poverty rates correlate with higher mobility rates. This population is significantly mobile: 30% of all kindergartens in 2011 – 2012 moved to a different school for first grade in 2012 – 2013.

<u>Pseudonyms</u>

The division's Pre-K program will be referred to with the pseudonym "New Day Pre-K" to distinguish it from Head Start, child care or private preschool providers. The New Day Pre-K program is described more fully in a succeeding section. The city and school division are stylized as New Day and New Day Schools to preserve anonymity.

New Day Pre-K Program

New Day preschool is a grant-funded early childhood program for four year olds. It is a free, full-day preschool run by the public city school system. Eighteen students are assigned to each classroom, with a certified teacher who is endorsed in early childhood education and an instructional assistant. All four centers provide the same curriculum, materials and teacher certification. Funds are contributed from the Title I program, from the State and from the school district's budget. Collaborative special education classrooms are provided at each site and ESL is available.

Income, home language, military affiliation, sibling participation, etc, do not factor into the choice of admitting a student into the program. All open slots are awarded based on academic need following a prescreening test. Students with the lowest scores are placed on the top of the ranking system, moving down the ranked list as scores rise. Most high performing students will ultimately be admitted because of vacancies. Children must be four years old by September 30th and a city resident to be admitted.

No student is formally "rejected," although their name may not come to the top of the list for selection until students with lower academic readiness scores are selected. A student's place in the ranking changes with every screening, as more students' scores are added to the total pool of applicants. Thus a student may be at the top of the selection list one week, but will be moved to second place as soon as another student takes the prescreening test and earns a lower score.

Nearly all high scoring four-year-olds are eventually admitted because of the large number of spots available in the New Day Pre-K centers. The smallest center has ten classrooms and serves 180 children. The largest center has thirty-four classrooms serving 612 preschoolers. During the current screening and admission window (2013), thirty students on the waiting list for the largest center could not be admitted as there were no vacancies remaining. However, all other centers had enough spots to accept every student, meaning that only thirty students were turned away in the entire city during the current enrollment cycle for the 2013 – 2014 academic year. Denial numbers are not available for the 2010 – 2011 year, although the total number of students turned away is small each year.

Screening appointments can be made online, by phone or at elementary and early childhood schools across the division. Following the screening appointment and the corresponding readiness screening test, parents are not informed about the position of their child in the ranking system (it fluctuates weekly based on incoming scores) or about their child's individual performance on the screening test. Upon acceptance, parents receive two letters informing them of their child's selection. Parents can then register their child at the school during a one week registration window, although they typically

will not lose their spot so long as they register within 2 – 3 weeks. Beyond that, the spot will be offered to the next child in line on the list as a result of registration no-shows. A high level of transience can cause considerable shifting in vacancies and registrations across the city.

Child requirements necessary for continued participation are communicated to parents in four unique venues: on the New Day Pre-K flyer, at the screening appointment, in the parent handbook and at the preschool orientation. These child requirements are: 1) to be completely toilet trained, allowing for accidents in the first month of school and allowing a maximum of three more accidents thereafter before removal from the program, 2) be picked up at the bus stop or at school (children may not walk home from school or a bus stop unaccompanied by a selected adult) with removal from the program on the fifth adult failure to meet the child at the bus stop or after school. Children are returned to school when an adult is not present for child pick-up at the bus stop. When these two requirements for continued child participation are not met, warning letters are sent home, parent conferences are held and toilet training advice is given, as the case may be, giving parents ample time to correct the issue before facing program removal.

Children are only admitted to the early childhood center in their zone and cannot attend or apply for a center outside their zone. The grant-funded nature of the program includes guidelines regarding zoning and attendance that are non-negotiable. Demographic changes in the census may cause zone changes to relieve wait lists or to fill vacancies.

Quality of New Day Pre-K

New Day Schools use structural indicators to suggest overall program quality. These structural features include: a limit of 18 children per classroom; a certified teacher in every room with a trained teaching assistant; early childhood certification for every teacher to support developmentally appropriate practice; a principal at every early childhood center and a shared standards-based curriculum across all centers based on the state preschool learning objectives. Total literacy growth from the beginning to the end of the year for children from each center is also considered to be an indicator of quality. There are instructional walk-throughs in the teacher evaluation process. Rating centers for quality by observation using a checklist, such as the use of environmental or classroom rating scales, is not conducted. Structural quality is the same across all five centers. The length of the instructional day, and the number of minutes for various parts of the day, is likewise the same across all centers. There is low teacher turnover in the New Day Pre-K program. For example, out of at least ninety-two certified Pre-K teachers, just two applied to transfer out of Pre-K during the 2012 – 2013 year. The teacher turnover rate at the centers has historically been low.

Daily Schedule in the New Day Pre-K Program

Instruction proceeds in terms of units based on standards and objectives set by the state for public preschool. The day is seven and a half hours long, which is the same across all centers. There is a thirty minute lunch. Children have a thirty minute free recess

block and a daily thirty minute structured physical education program. Children do not attend other resource classes (art, music, library, computer lab). Therefore, teachers do not have a planning block during the day. There is no daily snack or nap time. The day also includes two center rotation blocks: one for language arts, and one for open centers in math, science, drama and transportation.

New Day Pre-K Curriculum: 2010 – 2011

The language arts curriculum being used during the 2010 academic year was based on the Harcourt Trophies Pre-K Program, which came to the city in 2004. Math was taught for a few days a week. Math or science related centers were rotated but there was no formal math curriculum. At the time, the division's department of curriculum and instruction was not involved in the New Day Pre-K program. The curriculum was divided into a series of themes and units (See Table 1.1). A daily literacy lesson plan was prepared for teachers covering each day of the week, as Day 1, Day 2, etc. Center activities were suggested in the curriculum in the areas of literacy, writing, listening, math, science, art, dramatic play, manipulatives, water, sand and computers.

Table 1.1

List of Themes in the New Day Pre-K Curriculum in 2010

Working Together	Families	Growing Up	Fall / Harvest	Five Senses
Making Things	Foods	My Home	Winter	Neighborhood
Jobs People Do	Going Places	Pets / Backyard Creatures	On the Farm (Animals)	Zoo / Circus Animals
Weather / Spring	Wonderful Water	Ocean Animals		

Teacher Training in New Day Pre-K

Being part of the public school system, the teacher training schedule and approach is somewhat similar to that of K - 5 elementary school teachers. Lead teachers in each center act as a conduit between central office (new instructional approaches) and the other preschool teachers. Staff development days follow the same schedule as the school system. Professional development focuses on strategies to teach emergent literacy and math and the preparation and implementation of centers.

Social and Emotional Development and Behavior in New Day Pre-K

The division expects that the requirement that all teachers be endorsed in early childhood education supports the use of appropriate practices related to social and emotional development. The centers administer a preschool version of effective school-wide discipline (ESD) to reinforce and reward building-wide behavioral expectations. Individual classrooms use a color system to provide a visual cue to the child of his/her behavior, a leveling system that includes moving up one level above the starting position for good behavior. The system involves children's names moving up or down depending on their behavior. During the 2010 academic year, several social and personal skills were graded on the Pre-K report card, shown below (Table 1.2).

Table 1.2

Civics and Citizenship on the New Day Pre-K Report Card in 2010

1. Participates in creating classroom rules
2. Recognizes the need for & follows school rules
3. Listens to and follows one-step directions
4. States personal plans for learning center activities
5. Follows classroom routines
6. Cooperates with others in a joint activity & turn taking exchanges
7. Participates constructively in group situations
8. Identifies others' needs by helping them
9. Stays in an activity for an appropriate amount of time
10. Manages transitions
11. Handles classroom materials appropriately
12. Takes care of personal property
13. Demonstrates knowledge of personal information including first & last name, gender, age & birthday.
14. Participates in discussing and generating solutions to a class problem
15. Shares thoughts & opinions in a group setting
16. Copes with minor disappointments
17. Uses words and/or adults to resolve conflicts
18. Demonstrates respectful & polite vocabulary
19. Manages personal hygiene

Social (Citizenship) Skills on the First Grade Report Card

By the time children reach K – 3, the citizenship skills being graded on the report card have changed. At the end of each quarter in first grade, seven citizenship skills related to behavior are subjectively evaluated by the teacher and a grade is assigned to the child on the report card. The seven behavioral skills receiving a grade in first grade are: 1) Follows classroom / school rules, 2) Respects the rights of others, 3) Responds appropriately to authority, 4) Strives to achieve citizenship goals, 5) Takes responsibility

for learning, 6) Takes responsibility for own behavior, 7) Works cooperatively with others.

Research Questions

1. Does attendance in New Day Pre-K make a sustained difference in children's later literacy achievement in first grade?

2. Does attendance in New Day Pre-K make a sustained difference in children's social and emotional adjustment in first grade?

Design & Methods

<u>Grounding Methods and Design in the Literature</u>

The existing literature shows that about twenty studies have investigated the sustained effect of child outcomes into grade school as a result of attending state-provided pre-kindergarten administered by school districts. This narrow body of research has several identifiable characteristics, strengths and weaknesses that help inform and ground the design and methods chosen for this study.

The current Pre-K longitudinal effect literature struggles with a key weakness: the difficulty controlling for extraneous variables in the absence of a controlled trial (random assignment). Several studies provide questionable evidence of equivalence between the treatment (Pre-K) and control (no Pre-K) groups (Frede et al., 2009; Xiang &

Schweinhart, 2002) and lack a strong matching method. If the control group is more disadvantaged than the treatment group the longer-term findings cannot be attributed solely to Pre-K attendance, but to various other external factors. In the absence of random assignment, control and treatment cases can be matched to each other by representing all covariates as a single summed score assigned to each child, called the propensity score (Stuart & Rubin, 2004). Each child's covariate results are summed and a single score is assigned to the child, and children are then matched based on having identical or very similar propensity scores. This allows naturally occurring control and treatment groups, such as those found in educational treatment research, to be properly matched by pairing cases that have the same or nearly identical propensity scores. This method for balancing covariates among control and treatment groups (propensity score matching) has emerged as a current method of repute for much quantitative social science research when groups cannot be randomly assigned (Stuart, 2010). Cases that fall outside of the common support region are excluded from the sample. The remaining cases can be matched one to another, to form the control and treatment groups, with a significant degree of confidence that balance exists, so long as a large sample is employed with a large number of covariates in the model. While a range of statistical methods have been used in this literature, such as multi-level modeling and various regressions (Sylva, 2004; Peisner-Feinburg, 2008), only one study used propensity score matching to approximate the balance attained by random assignment (Magnuson, 2007a).

The use of a matching method, in particular the propensity score matching (PSM) technique, brings rigor to the design. The balance PSM creates among groups makes analysis more reliable. Tests of significance, such as a one-way analysis of variance, are

only trustworthy when the groups have been confidently balanced beforehand. To use PSM appropriately, this study needed a large sample and a large number of covariates, conditions that were met by the existing data. After reliable matching on the propensity score, given both a large number of covariates and sample, an established test of significance could be used to analyze possible causal comparisons.

The longitudinal Pre-K effect literature primarily focuses on reading and math achievement, which is important, but only a few studies include measures of both social and academic achievement (Sylva, 2012; Berlinski, 2009; Peisner-Fiensberg, 2008; Xiang, 2002). Measuring a child's later social and emotional adjustment as it potentially relates to having attended Pre-K remains an open question given the lack of research on this particular question. The current study design reflected the conceptual understanding that later school quality mediates longitudinal effects. For example, when Magnuson et al., (2007a) introduced a measure of later school quality control into their design they discovered significant sleeper effects of Pre-K attendance re-appearing in third grade which were not discovered within the first grade data, reported separately (Magnuson et al., 2007b). The design of the current study reflected strengths in the literature, such as the use of propensity score matching and analysis of later school quality in relation to outcomes. The current study also balanced outcome measures to be both realistic for program evaluation (literacy) and holistic for child development (social and behavioral). Using such design and methods, the study contributes to the discourse on providing large-scale penetration of district Pre-K, while also addressing a historical concern with fading effects (Lee & Loeb, 1995).

Design

A quantitative, observational study with an ex post facto (causal comparative) design was used to investigate the later impacts of New Day Pre-K attendance in a large urban school district. The study used existing archive data to investigate the research questions.

Matching

A large number of covariates and a large sample are well suited to the use of propensity score matching to create approximately balanced control (no formal or institutional preschool) and treatment (attended the district Pre-K) groups. Matching was achieved using the optimal matching technique on a fixed ratio, in which one control case was matched to five treatment cases because this study had a significantly larger number of treatment cases than control cases. Commonly, one-to-one matching would be used whereby each case with the nearest propensity score is matched in nearest neighbor, or greedy, matching. However, the optimal matching method makes the best global decision compared to other matching approaches, such as greedy matching, by minimizing the average absolute propensity score differences (Gu & Rosenbaum, 1993). One-to-many matching allowed one control case to be matched to up to five treatment cases, thus preserving as much of the treatment sample as possible.

Analysis

Following a demographic covariate balance check after matching, the analysis

proceeded in two steps using inferential statistics.

1. Independent-samples t test will be used to determine if there is a significant

    difference between the two groups on various continuous dependent variables.

2. P-value scores from the t tests will be converted to effect size scores to calculate

    how strong an effect Pre-K attendance may have on each dependent variable.

Measures

Literacy was measured using continuous variables (DRA2 independent reading

levels and PALS literacy scores). Both of these measures are standard tests of reading

achievement in districts across the state. Behavior was measured, as a proxy for social

and emotional development, with a continuous scaled dummy variable representing a

sum score of seven student citizenship skills taken from report cards.

Data collection

Demographic data (age, race, income, zip code, school name, school quality,

special education or disability, retention and mobility) were collected from the school

district in archived format and merged into a single file. Student preschool experience

data were gathered from the director of Title I programs overseeing the Pre-K program. Reading tests and report card data were collected from the director of research at the district.

Limitations

A causal comparative study faces limitations versus an experimental design. This study relied on propensity score matching in the absence of random assignment to approximate group equivalence. Selection bias may be present due to uncontrolled confounding. However, a large sample size and large number of covariates were represented in the model.

While two measures of literacy were included using scored instruments with appropriate reliability and validity, subjectivity was present in teacher report card data used to measure student behavior. However, it is very unlikely that a poorly behaving child would receive satisfactory marks from a teacher across all seven citizenship skills on his or her report card. Rather than comparing students to each other based on any particular citizenship skill, which was overly subjective, a continuous scaled index variable was created to represent a child's overall number of S's, N's or U's earned across the seven citizenship skills. This created a single scaled sum score suitable to analysis while avoiding some of the subjectivity associated with comparing one skill to another.

Results may not generalize to the region or country as a whole. However, this district is a model of Pre-K access, having one of the highest percentages of students

enrolled in Pre-K. The district serves a racially and economically mixed population in an urban environment that is similar to many other cities in the United States. The causal comparative nature of the study, with group equivalence by propensity score matching, and a large sample size, all helped to improve the generalizability of the findings. Also, a more conservative effect size was reported. Reporting a conservative effect size helped account for some additional variance, found in other regions and areas, to support generalizability.

Significance

Very few studies have examined the longer-term impact of public Pre-K (as opposed to Head Start or other preschool) in the United States (n=8) with just one published study found at the school district level (Valenti, 2009). Conditions under which Pre-K effects will be sustained are not well analyzed in this limited literature, with all but one study (Magnusun et. al., 2007b) neglecting to include a measure of later school quality in the analysis. The confounding nature of student mobility and later school quality have not been simultaneously controlled in any study on the longitudinal effect of Pre-K attendance into first grade (chapter 2).

This study has important implications for practice on how to best align public PreK with K-12 education, supporting current efforts to create "a PreK – 3 model" (Pianta, 2009). The study is significant as a potential model of public Pre-K access (because just 1 % of applicants are turned away) that can potentially be adopted by other districts, and to help districts learn how to "create a more coherent and uniform platform"

(Pianta, 2009). Expansion of public Pre-K to all children is a national topic of discussion and this program may be useful as a case study in providing nearly universal Pre-K access in a mixed and urban setting, in a state that lacks a mandate for universal Pre-K. As such, other states and districts may be interested in the example and findings.

The study is significant because it will help to evaluate and improve a unique and specific district program. This Pre-K program had not been re-evaluated beyond school readiness since the new state preschool standards took effect in 2005. Findings can help the school district, and future researches, to better understand how to target intervention for at-risk children across the first few years of schooling.

This research contributes to the wider early childhood education field regarding PK – 3 alignment, provides further evidence regarding the sustainability of Pre-K literacy impacts into first grade, provides new evidence about the relationship between behavioral outcomes by first grade and Pre-K attendance, and enhances the focus on later school quality in relation to evaluating Pre-K using longitudinal designs.

Operational Terms

State-funded preschool, prekindergarten or "Pre-K" – publically funded preschool administered by a state government and implemented by public school districts, as an addition to the K – 12 school system, but usually limited to children with one or more risk factors such as income and academic need. Pre-K is often funded with Title I federal money and as such is often considered part of Title I programs. There are limited spots available in most states, and often with low-income criteria and wait lists for acceptance.

This study uses the term Pre-K as a publically funded preschool program that is not the United States Head Start Program.

Head Start – as distinct from *Pre-K*, Head Start is a federally funded public preschool program for three and four year olds. Many children move between Head Start and Pre-K, and some school districts are involved in the administration of Head Start as a component of their preschool program(s) or administer them in tandem to build a more complete early childhood education program.

School Readiness – early literacy, numeracy and social skills at kindergarten entry.

K – 3 – kindergarten through grade three.

PK – 3 – preschool through grade three.

Early Childhood Education – infancy through grade three, ages birth – 8.

CHAPTER 2

LITERATURE REVIEW

Organization

This review begins with a brief overview of the historical context behind pre-kindergarten education in the United States. The current environment of Pre-K, including current trends and characteristics, is then presented. Indicators of higher quality preschool follow this introductory information. The literature review is then presented using an original conceptual framework. A critical literature review of the methodological approaches to longitudinal Pre-K evaluation studies, including the literature on sustaining the gains of Pre-K attendance beyond school readiness, is included. Finally, suggestions for a new study based on the existing literature are presented.

Historical Context

The flagship program for providing early intervention services to disadvantaged children by the federal government is Head Start, governed by the Head Start Act of 1981 and reauthorized in 2007 (Head Start Act, 2007). Head Start has served more than 30 million children since 1965 and currently serves more than 900,000 children and their parents who are living near or below the federal poverty line (Administration of Children and Families, 2013).

While Head Start continued to slowly expand in slots and funding from the late 1960's to the early 2000's, several social, economic and scientific trends (post 1960) converged to precipitate the creation of large-scale pre-kindergarten administered by state departments of education and school districts.

For one, the landscape of gender and labor changed. From 1950 to 2000 the number of women working grew from 16 to 66 million to comprise 47% of the labor force (U.S. Bureau of Labor Statistics, 2002). The prevalence of divorce and single parenting increased. These changes created a significant need for child care services, while poverty rates also increased, as described below.

The steady increase in the number of families and children living in poverty necessitated the search for expanded intervention programs. A brief overview of under-six child poverty rates illustrates this trend. While the child poverty rate declined from 1959 – 1968, it has been on an upward trend ever since, moving from a low of 15% in 1969 to over 25% by 1993 (U. S. Bureau of the Census, 2008). Under-age-six poverty fell to 18% by 1999, but was back up to 20% by 2004. By 2008, families and children had entered a recession and a prolonged economic contraction. The under-six poverty rate hit 25% again by 2010 (National Center for Children in Poverty, 2012). Funding cuts to benefits and social services persist today.

To further complicate the situation, the minimum cost of living is much higher than the official poverty line (Cauthen & Fass, 2008). In 2008, the minimum income needed to cover basic expenses for a family of four people ($64,000 in Los Angeles to $42,000 in Jackson, MI), was twice the federal poverty limit. If poverty is recognized as not being able to meet ones basic living expenses, now termed "low-income," then the

rate of child poverty soars to 48%. This is the picture today, where about 1 in 2 children in the United States live in either poverty or low income families (National Center for Children in Poverty, 2012).

The effect of deprivations related to poverty on child development is well-known (Knudsen, Heckman, Cameron, & Shonkoff, 2006) in terms of poorer health, social and emotional risk, and reduced cognitive development compared to middle-class peers. As the need and demand for early education services was rising, a corresponding recognition among the scientific community and the public about the significance of early childhood development and the value of early experiences emerged. By the 1990's, calls for universal or expanded preschool, reminiscent of Kindergarten expansion, were underway to both improve cognitive, social and health outcomes while assisting low income families and children.

The movement in the 1990's to provide public preschool education connected to K – 12 schooling gave rise to the rapid growth of pre-kindergarten programs in states around the county. The first efforts to create truly state-wide Pre-K programs occurred in Georgia and Oklahoma in the early 1990's (Southern Education Foundation, 2007). Georgia announced that voluntary Pre-K was available to all four-year-olds regardless of income in 1995, the first state in the history of the country to provide universal preschool. Oklahoma followed suit shortly thereafter.

Current Scope of Pre-K

The National Institute of Early Education Research (NIEER) publishes an annual report called the State Preschool Yearbook. The most recent edition of the Preschool Yearbook thoroughly describes the current state of public pre-kindergarten, the quality of state programs, growth and funding, and analysis of every state's Pre-K picture (Barnett, Carolan, Fitzgerald, & Squires, 2011). The report shows that 1.3 million children attend state Pre-K programs, 28% of all four year olds. Head Start, meanwhile, serves another 11% of all four year olds. Twenty-nine percent of four year olds were in private care and 26% stayed at home. The percentage of four year olds enrolled in state Pre-K has risen by approximately two percent per year from 14% in 2002 to 28% in 2011.

According to the 2011 Preschool Yearbook, funding of Pre-K has not kept pace with enrollment growth. Average state spending per child decreased from about $4900 in 2002 to $4200 in 2011. Thirty-nine states operate Pre-K programs of various sizes. Arizona eliminated funding for its program, bringing to 11 the number of states with no state Pre-K program. The concern is that current funding is too low, in all but 12 of the 39 states with state Pre-K, to pay for ten basic quality standard benchmarks (described in the next section). As such, 43% of children in Pre-K, or more than half a million children, are served by programs that meet fewer than half of the structural quality benchmarks. Enrollment growth and the ability to pay for various quality measures are largely shaped by state legislatures through the budget allocated to Pre-K operation or expansion. The plateau in national enrollment growth in recent years has not been for a lack of demand, but a lack of access due to constrained budgets.

Indicators and Measures of Quality

Negative consequences await children after low-quality early education. For example, if they do not reach reading proficiency by third grade, 30% of poor African-American and Hispanic children will fail to graduate from high school later on (Hernandez & Foundation, 2012). One great year of early education, followed by several mediocre years, does less good (Nelson et al., 2003).

However, there are perspectives on quality at the preschool level that should be reviewed when determining what a baseline of Pre-K quality might mean. Preschool quality is judged from numerous perspectives, such as structural quality (NIEER), process quality (ECERS), teacher effectiveness (CLASS), theoretical orientations (constructivist, etc), developmentally appropriate practice, academic orientation, and even economic return on investment. Professional development, intentional instruction and an early literacy focus are strong predictors of kindergarten outcomes to use as measures of preschool quality (Williams, et al. 2012). These various perspectives towards quality all aim at the goal of enhancing child development, which is confirmed by measuring particular child outcomes.

Structural quality indicators include class size, teacher-child ratios, teacher qualifications, hours of service, meals, health support services, etc. An example of widespread use of structural indicators to determine overall quality of state Pre-K programs is the National Quality Standards Benchmarks checklist used by NIEER. The National Institute of Early Education Research (NIEER) at Rutgers University publishes

an annual survey to evaluate state's Pre-K programs on ten indicators of quality. The benchmarks include: 1) a comprehensive standards-based program (to classify the program as academic or not sufficiently academic in nature), 2) bachelors degree by teacher, 3) teacher specialization in Pre-K, 4) assistant teacher certification in early childhood, 5) teacher in-service hours, 6) class size of 20 or less, 7) staff-child ratio of 1:10 or less, 8) vision, hearing and health screening and referral and a support service, 9) meals and 10) site visits to monitor quality. Earning ten out of ten points is the highest quality ranking given to state Pre-K programs based on this checklist (Barnett et al., 2011).

Process quality, distinct from structural quality, refers to the child's lived experience in the preschool setting, including activities, materials and interactions with others (Phillipsen, Burchinal, Howes, & Cryer, 1997). The child's lived experience (process quality) is more predictive of later outcomes than structural quality (Whitebrook, 1989). A common instrument used to measure process quality, which is observed quality on-site, is the Early Childhood Environmental Rating Scale - Revised (ECERS - R) which assesses basic care of children, physical environment, curriculum, schedule, interactions, program structure, parent and staff needs (Clifford & Reszka, 2010).

Teacher effectiveness is an indicator of Pre-K quality. Teacher effectiveness includes instructional quality, responsive relationships, classroom organization and supportive feedback, among other factors. An example of measuring teacher effectiveness, particularly the quality of various aspects of adult-child interactions, is the *Classroom Assessment and Scoring System* (CLASS). The CLASS measures adult-child interactions across three main domains: Emotional Climate, Management, and

Instructional Support (Pianta, La Paro, & Hamre, 2008). A responsive adult-child relationship and intentional social and emotional development are part of high quality preschool.

In addition to adult-child interactions, the effective early childhood teacher understands and implements developmentally appropriate practice by matching learning experiences, teaching methods and types of activities to the appropriate needs of each individual child and to specific groups of children (Bredekamp & Copple, 1997). Education is developmentally appropriate when it is matched to the learner's ability, meeting children where they are, while setting challenging and achievable goals (Copple & Bredekamp, 2008).

Quality is also related to the application of particular theoretical orientations, or philosophies, associated with early childhood education. A particular theory (or theories) of child development, such as constructivism (Piaget) and social-cultural theory (Vygotsky), may frame quality as following child interests, encouraging extended play, emphasizing inquiry, exploration and projects, creating authentic and meaningful experiences, and other approaches and activities resulting from respected theories of child development.

Regardless of perspective and orientation, quality is eventually judged based on child outcomes, as well as the provision of family and child support services. Child outcomes are most commonly related to physical, affective and cognitive child development. As formal schooling concerns itself primarily with cognitive development, there is a heightened focus on school readiness in terms of early numeracy and literacy, as well as social and emotional adjustment for school success. Children's later

performance gains (impacts into grade school) are also being explored as a potential indicator of Pre-K quality and as a means to improve PK – 3 education as a whole.

Quality in the Context of Typical Pre-K Versus Intensive Interventions

The expectation for longer-term impacts related to preschool quality requires further research. For example, recent results from England show that high quality preschool can lead to extended grade-school impacts (Sylva, 2012), but scant research exists to confirm these findings for children in the United States in relation to typical Pre-K as it is emerging around the country.

Some arguments in favor of expanding preschool use the results of intensive programs as evidence. However, typical Pre-K offered by school districts is different from the intensive interventions designed specifically for high risk students, such as the Abecedarian preschool program in North Carolina or the Child-Parent Centers in Chicago, which provide wrap-around and comprehensive services to young children and their families over several consecutive years.

Intensive programs are recognized as standard-bearers of a particular form of "high quality" because of their ability to make long-term impacts on children's well-being, including benefits to children's later health, improved high school graduation and decreased incarceration rates (Reynolds, 2010). However, it is an assumption to propose that typical Pre-K impacts will equal those of intensive preschool programs. Yet, reducing the expectation for Pre-K benefits to school readiness alone may also not explain or reveal the full value provided by Pre-K. The question of how various degrees

of Pre-K quality relate to later outcomes requires more study to reach a realistic consensus on the true effectiveness of Pre-K for various groups of children.

The Literature Review

Introduction to Longitudinal Pre-K Evaluation Research

State Pre-K evaluations most typically report school readiness outcomes at kindergarten entry (Wong, Cook, Barnett, & Jung, 2008). Longitudinal studies of Pre-K outcomes beyond kindergarten are expanding as well (Pianta, Barnett, Burchinal, & Thornburg, 2009). Previous evaluations are being improved upon by using more advanced statistics and by lengthening duration (Huang, Invernizzi, & Drake, 2011).

Barriers to data collection are complicating these efforts. Hernandez (2012) reports the disconnected Pre-K and K-12 systems, various forms of data collection, inadequate computer database systems, absence of common assessments, unconnected databases, student transience, and confidentiality concerns as the greatest difficulties hindering longitudinal Pre-K research. Pre-K from state to state is diverse, affecting replication efforts: states do not have uniform standards, assessments or curricula, and definitions and levels of quality vary significantly (Barnett et al., 2011). Regardless, state departments of education must evaluate Pre-K programs for two reasons: (a) to justify continued funding, and (b) to reach higher quality because it is effective (Mashburn et al., 2008).

The Necessity for the Current Review

The review presented in the following sections is unique to the literature by focusing on Pre-K exclusively and its outcomes beyond school readiness. Other reviews of public preschool encompass both Head Start and Pre-K (Barnett & Frede, 2010) or mixed preschool options, while state Pre-K evaluations usually look only at school readiness indicators (Wong et al., 2008). Typical Pre-K programs are also distinct from comprehensive interventions such as the Abecedarian, Child-Parent Centers or High/Scope Perry programs.

Questions About Pre-K Evaluation

Initial questions about Pre-K evaluation include: 1) should results of Pre-K programs be considered beyond school readiness? If so, 2) how will fade-out effects be addressed (Currie & Thomas, 2000)? 3) To what extent can Pre-K compare to comprehensive preschool programs (Temple et al., 2010)? 4) Can quality Pre-K improve school readiness and provide later impacts vital for vulnerable children? Longitudinal research will help to clarify these questions.

Conceptual Framework

The purpose of this review is to critique longitudinal Pre-K impact studies that report outcomes into elementary school. While the official purpose of Pre-K is school readiness (Gilliam & Zigler, 2001), longer-term research helps to align PK-3 (Reynolds et al., 2010), to ensure quality through consecutive grades, address fading effects and to search for ways to sustain the gains made in Pre-K years.

Seeking sustainable gains

Lacking sustained impacts on broader measures, Pre-K serves a functional role as another grade of public schooling, similar to kindergarten expansion decades before, with a narrower focus on cognitive gains for the short term (Gormley Jr, Gayer, Phillips, & Dawson, 2005).

Balanced child development in measuring outcome domains

Pre-K programs are risking the loss of a social-emotional emphasis and holistic potential that is effective in comprehensive interventions (Reynolds et al., 2010). Holistic child development is foundational to early childhood education. It is only natural, then, that attention to holistic development should remain at the center of a conceptual framework for longitudinal research. Therefore, impacts across various developmental

domains must be investigated, particularly the balance between affective and cognitive development.

Data limitations, grant or agency requirements, and the orientation of the researcher often determine the diversity of outcome domains that are measured in longitudinal research on public preschool. Such measures may be cognitive, social-emotional, physical or health related, or based on specific outcomes such as retention, delinquency, adolescent or even later adult impacts. When the majority of research measures cognitive development but not social and emotional development, for example, researchers may unintentionally contribute to the view that what is evaluated is the most important aspect. It is easier for practitioners to maintain a developmentally appropriate and holistic approach to education when evaluators measure a balance of outcome domains (Falk, 2012).

Relationships between fading effects, teacher effectiveness and later school quality

While balanced and sustained Pre-K impacts are sought, it is necessary to consider fading effects: the observation that preschool benefits fade over time (Lee & Loeb, 1995). Several reasons for fading effects are considered in the literature. Non-preschool children may catch-up to their peers (Magnuson, Ruhm, & Waldfogel, 2007a), perhaps after receiving more help from the teacher if they are initially performing below the Pre-K group, or by attending better schools. Conversely, higher students may regress after receiving less teacher attention in an era of accountability to meet minimum benchmarks (Loveless, Parkas, & Duffett, 2008). Furthermore, Pre-K children tend to be

from low-income populations due to enrollment criteria set by many states and districts, meaning they often live in lower income neighborhoods. They frequently go on to the lowest-performing schools in the district which fail to sustain the preschool gains (Currie & Thomas, 2000).

To illustrate the issue related to later school quality and teacher effectiveness, if a teacher at a mediocre school receives a high-performing student but cannot maintain that child's high achievement, the concern is that the teacher allowed the child to regress to the mean (fading effects), perhaps because attention is given to the most under-achieving students. An incorrect culture that minimum benchmarks are sufficient for all children, and that high achievers do not need to remain high, implies that the profession is neglecting the importance of keeping a child on his or her potential learning trajectory. Failure to sustain a child's learning trajectory would not meet the definition of "developmentally appropriate practice" which requires setting challenging but achievable goals for each young student based on his or her unique achievement (Copple & Bredekamp, 2008). This loss of focus on accelerating children as high as they can go, regardless of their starting point, could be due to the entire mechanism of tying minimum standard benchmarks to teacher, school, district and state evaluations (Loveless et al., 2008).

Changing teacher evaluation: from minimum benchmarks to equity

State teacher evaluations are now decoupling minimum standard benchmark scores from teacher evaluations in favor of a more equitable system, whereby each child

should make a full year's worth of growth, according to ability (a full years worth of growth for a slow learner versus a fast learner are not the same), regardless of the child's starting point as high, low or average. Such a system places responsibility on the teacher to ensure that every child makes expected gains, rather than neglecting top students, giving only basic attention to average students, or placing a majority of effort on the lowest achievers.

Addressing fading effects and later school quality in future longitudinal research

A study design to control fading or growing effects due to school quality would follow Pre-K children into grade school and then compare their performance to Pre-K children who went on to mediocre schools. Such research would reveal both the future potential of vulnerable children across school quality settings while also providing insight into the ability of Pre-K to make sustainable gains. High, mediocre or low quality schooling would need to be defined.

To address fading effects using these conceptual parameters, longitudinal designs need to control for grade school quality, use continuous variables for achievement (rather than dichotomous measures), and use strong matching designs to control for a large number of covariates. (This is explored later). Studies should encompass several years of grade school to monitor learning trajectories over time, within and between the control and treatment groups.

Pre-kindergarten curricula

The exact curricula used by different Pre-K programs should be reported and examined because some curricula seem to have little to no effect, and even detrimental effects, while others have proven to be effective on a range of outcome domains (What Works Clearinghouse, 2012).

Reporting Pre-K quality

This review is informed by the concept that existing preschool quality must be considered before reporting longitudinal effects. Later child outcomes are different depending on high, medium or low quality preschool received (Sylva et al., 2011). Such reporting of original Pre-K quality should go beyond structural quality benchmarks, discussed earlier, and include classroom climate and teacher-child interactions because these processes correlate to later achievement (Curby et al., 2009).

Methods

A systematic search for empirical literature was conducted to find Pre-K impact studies beyond school readiness. The following databases were searched: ERIC, EBSCO, JSTOR, Education Full Text; PsycInfo, Google Scholar, SAGE, and ScienceDirect using combinations of key search terms: (a) pre-kindergarten, preschool, pre-primary, pre-k; and (b) effect, impact, outcome; and (c) longitudinal, grade school,

long-term, first grade, second grade, third grade; and (d) school quality, sustained, fading effects, and duration. Direct searches of research institutes and governmental websites yielded more studies. Citations and references in studies were combed for additional results.

The following criteria were set for inclusion: (a) longitudinal evaluations about the effect of attendance in state-funded prekindergarten programs, such as district provided preschool; and (b) published in peer-reviewed journals, produced by government or research institutions, or contracted evaluations of state programs (dissertations were excluded to focus on professional evaluations); and (c) studies needed to be published after 2000 so as not to overlap with significant meta-analysis and summaries of the research to that point (Gilliam & Zigler, 2001; Camilli et al., 2010); and (d) longitudinal duration had to extend beyond school readiness, to the end of kindergarten at a minimum, preferably into first grade or beyond, to distinguish the current review from the more abundant school readiness literature; and (e) methodology was included in selection criteria: studies were included if they had experimental or quasi-experimental designs.

Follow-up studies of foundational comprehensive programs, such as Abecedarian, Child-Parent Centers and the High/Scope Perry programs, were excluded from this review in order to focus on typical large-scale Pre-K as it exists more commonly throughout the 39 states offering programs (Barnett et al., 2011), and internationally.

A search in July and August, 2012, yielded a total of 18 empirical studies for initial inclusion. After further investigation, one final report duplicated a later publication, and one study did not distinguish the child care providers. Therefore, a total

of 16 evaluations were selected for analysis. These studies were then classified by location, scale, population, duration into grade school, methodology, effect and limitations. This initial classification was followed by an analysis based on the following criteria: (a) preschool quality measurement (b) later school quality control, (c) reasons for fading effects, (d) curriculum reported, (e) outcome domains, and (f) challenges (design, measures, statistics and results). The degree to which a study addressed these latter criteria, as grounded in the conceptual framework presented above, informed the critique.

Results

Longitudinal studies evaluating the effect of state-funded pre-kindergarten from 2000 to 2012 are included in this critical review. Tables 2.1 and 2.2, at the end of the results section, present an overview of the findings.

Summary of locations, programs and duration

Ten unique locations were evaluated in all: three in other countries and seven in the United States. All of the studies will be critiqued against the conceptual framework and criteria already described. In addition to these studies, Waldfogel & Zhai (2008) investigated seven OECD countries (Japan, Norway, Netherlands, Australia, New Zealand, UK, USA) in a group study on the effect of money spent for Pre-K on later math and science test scores in fourth grade. England, Uruguay and Argentina were each represented by a unique longitudinal study. Studies in the United States took place in

Virginia, the northeast (one school district), New Jersey, North Carolina, Arkansas,

Michigan and nationally through two dataset studies. Sample sizes ranged from a few

hundred (school district study) to 186,000 (adoption of universal preschool in Argentina)

but typically fell in the range of 500 – 10,000. Cohorts began prekindergarten as early as

1997 in Michigan (Xiang et al., 2000) and as late as 2007 in Virginia (Huang et al.,

2011).

The longitudinal duration of the studies ranged from examining the effect of Pre-

K by the end of kindergarten (Frede, Jung, Barnett, Lamy, & Figueras, 2007) to the end

of seventh grade (test results) and eighth grade on retention or other factors (Maloffeva,

Daniel-Echols, & Xiang, 2007). Two studies examined longitudinal Pre-K effects into

middle school (Sylva et al., 2011; Maloffeva et al., 2007) and other studies focused on

effects into first grade (Hustedt, Barnett, & Jung, 2008).

Synthesis and critique: domains, measures, curriculum, Pre-K quality, later school

quality, fading effects, limitations and design

The outcome domains almost always included language and literacy in some

combination (15 out of 16). Math effects were measured by 12 of the studies. Social and

emotional domains, including behavior and self-control, were measured in just 5 of 16

studies. Five studies measured retention. The Michigan studies included other measures

for attendance, special education placement, extra services provided and later advanced

course selection. Later child health or motor development in grade school, as a

correlation with Pre-K attendance, was absent from this literature, although it has been

studied separately (D'Onise, Lynch, Sawyer, & McDermott, 2010). The guiding principle that child development is best measured using a balance of outcome domains, to include affective measures, was not well represented in the studies. Just three locations (19%) included measures for both cognitive and affective domains. These were England (Sylva et al., 2011), Argentina (Berlinski, Galiani, & Gertler, 2009) and the two Michigan studies in the United States(Xiang & Schweinhart, 2002).

The particular instrument(s) used to gather results changed as duration lengthened because of the nature of testing. In end-of-K and first grade studies, common measures included the PPVT and the Woodcock Johnson-III including math and reading subtests. Social and emotional data were mostly taken by teacher surveys and not by researcher observation. A strength of two studies was the use of one-on-one testing of children to gather latent ability scores which were later transformed (Peisner-Feinberg, Elander, & Maris, 2008; Magnuson et al., 2007a). This is important given the theory that dichotomous measures of child ability (passing / not passing), many standardized tests, and tests with a ceiling score (learning all letter sounds), fail to capture the true value of a child's skill or knowledge level on any given subject. Latent ability tests were not administered as measures in any studies beyond first grade.

Part of the conceptual framework related to preschool quality is the use of an evidence-based curriculum, one that has been tested in a variety of settings and found to be effective. In theory, researchers should check the preschool curriculum against existing empirical evidence, or test it directly, when conducting these evaluations. Just 3 of 16 studies made any mention of the curriculum used in Pre-K. Only Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart (2004) named the exact Pre-K curriculum

(Foundation Stage Curriculum). A few other studies listed some of the recommended curricula from which the sites had the option to choose. England was the only location to analyze both adherence to the curriculum and its quality, by comparing classroom observation scores, across a range of domains, to the curricular domains. Sylva (2004) reported the curriculum in England to be effective, that it was being adhered to, and concluded that adherence to the curriculum in Pre-K was still exerting a positive impact on children at age six.

To conduct meaningful evaluative research on the effect of Pre-K on later child achievement, it is valuable to have some measure of the quality of the Pre-K program at the start. As longitudinal studies extend their results into later grade school years, the need to control for school quality is extended, as some children go to a high performing school, broadly determined by test scores, while other children may go to a mediocre school. This difference in both Pre-K and later school quality causes confusion when attempting to interpret the results of the Pre-K experience. Controlling for quality involves more than controlling for school-level effects in hierarchical regression.

Only 38% of the studies (6/16) had a measure for Pre-K quality. This consisted of observational data using an instrument, such as the ECERS-R, to record scores for quality in various domains. One other location had no observed measure for Pre-K quality, but did mention characteristics of the program across the region that met pre-determined standards of quality such as class size, teacher-child ratio and teacher qualifications (E. C. Frede et al., 2009; E. Frede et al., 2007). Programs using national or state datasets were the least likely to include a measure of Pre-K quality, but they could have done so had they randomly selected a subsample of sites for observation to compare to the dataset.

Simply having a measure for quality does not mean that a particular study controlled for this variable in a statistical model. It appears that only the England study deliberately stratified Pre-K quality scores, in the design, in order to correlate specific quality profiles with longitudinal child outcomes (Sylva et al., 2011).

In England, Sylva (2011) created three Pre-K quality profiles (low, medium, high) using observed ECERS-R and ECERS-E scores. The ECERS-E measures quality of curriculum. She then tested child outcomes at age 11 against the low, medium and high quality groups and found significant differences for children which varied depending on original Pre-K quality. As the only study to stratify quality and report separate child results, it is worth reviewing the findings: (a) the low-quality Pre-K treatment group, versus the no preschool "home" control group, had identical scores at age 11 on math and English test scores; and (b) medium quality preschool children scored significantly higher, and (c) high quality preschool children scored the highest of all groups at age 11. Any quality of preschool made a positive difference on prosocial behavior at age 11. However, self-regulation scores were only significantly improved if the child attended a medium or high-quality Pre-K. Finally, only high-quality Pre-K significantly reduced child hyperactivity. Hyperactivity was measured using the Adaptive Social Behavior Inventory (ASBI) at age three as a baseline, followed by teacher rating scales using the Strengths and Difficulties Questionnaire. Hyperactivity was found to be one of the four underlying dimensions of child social behavior confirmed by factor analysis in the study.

Closely related to controlling for Pre-K quality is controlling for later school quality, to bring meaning to the interpretation of results from longitudinal studies, particularly for studies extending beyond the first one or two years of grade school. In

this case, just 19% (3/16) of studies had any measure of later school quality. Two of them used a variable for school quality in their models to test the effect on Pre-K longitudinal outcomes. The North Carolina study (Peisner-Feinberg et al., 2008) evaluated later kindergarten classroom quality using the ECERS-R, ELLCO, CIS (interactions) and the AEEC (cohort 2) instruments but only reported results as operational data: that the Pre-K classrooms were of higher quality than kindergarten rooms. In the evaluation on the uptake of universal preschool in Uruguay, Aguilar (2012) reported a "strong effect" on performance at year six (fifth grade) due to one variable of later school quality: the average rate of grade retention for the school over the entire study time (1999 – 2004). When the authors introduced this variable of school quality to their model they found a strong effect on fifth grade performance. (Magnuson et al., 2007a) also measured for school quality in their study on the persistence of Pre-K effects by third grade. In this case, the researchers created a score for quality by (a) the average amount of daily reading instruction (less than 90 minutes was considered low quality), and (b) class size (21 students or more was considered low quality). Results suggested that maintaining the advantage from Pre-K, versus other kids catching up, is a function of later school quality.

How well do the 16 evaluative studies analyze fading effects? In total, six of the studies analyzed fading effects in relation to their outcomes. Admittedly, it is more important to talk about fading effects when results are insignificant. Three evaluations reported insignificant academic results on child outcomes during the final year of their study. Of these three studies, two of them engaged in discussion on the topic (Magnuson, Ruhm, & Waldfogel, 2007b; Huang et al., 2011). The Virginia study suggested that for

Black, Hispanic, and students with disabilities, some benefits persisted to the end of first grade, such as retention reduction, but overall benefits were greatest at kindergarten entry. Magnuson et al. (2007b) found that literacy and math scores were not sustained through first grade but benefits did persist more for at-risk children. In a follow-up to this same study, the effect of Pre-K on literacy and math was found to be significant by third grade, suggesting that sleeper effects existed during the first grade time period. Finally, Maloffeva et al. (2007) was the lone study to find both no academic effects during the last testing window (seventh grade) and yet not discuss fading effects. This would have been an interesting discussion because the previous report tracking the same cohort (Xiang & Schweinhart, 2002) found significantly positive test results in literacy and math by fourth grade in a sample size of about 400 children. The change from significant results to insignificant results between fourth and seventh grade in Michigan would make for an interesting follow-up study, however, groups were not properly matched at the outset of Pre-K and selection bias would make true analysis difficult. Also, the fact that the Michigan study tracked children for so long into grade school compounds the issue of possible school quality influences, for which the researchers did not control.

The most robust design to address fading effects was the (Magnuson et al., 2007a) study because it controlled for later school quality, used latent ability test scores and an appropriate matching method (propensity score matching) to control bias between the treatment and control groups. The duration of the study was over several years (third grade) of early grade school to track child trajectories. The result of this robust design was the discovery of significant effects "re-appearing" in third grade. Both the finding and the design are promising signs for future researchers.

Design, statistics and limitations in the literature: strengths and challenges

All studies were quasi-experimental in design. They are correlational in nature, rather than causal. However, a matched pair design that controls a large set of matching covariates, with a larger sample, may be able to replicate the results of causal experiments (Stuart & Rubin, 2004). No study included a true experiment using random assignment to control and treatment groups. A random control trial is currently underway at the Peabody Research Institute in Tennessee which will be reporting the effect of Pre-K through third grade in that state, as mentioned earlier (Lipsey, Farran, Bilbrey, Hofer, & Dong, 2011).

For the present group of 16 studies, the strongest designs used propensity score matching to control covariates, reducing selection bias to approximate an experiment, as well as having very low attrition rates (three percent) and using dummy variables for missing data (Magnuson et al., 2007b). Weaker designs did not use multiple imputation for missing data, nor empirically test their attrition effects, nor use a matched pairing between treatment and control groups (Frede et al., 2009). The Abbott (New Jersey) studies lacked true matching. Using the same sample in both studies, the treatment group had 753 students and the control group had just 248 students. While the regression results may still be valid, one-to-many matching on a propensity score would have improved covariate balance between groups, strengthening the statistical results.

Another example of possible unequal control and treatment groups was found in Michigan where the criteria for Pre-K entry was having any two risk factors in the family (such as low income and a single parent family) while the control group was only

required to have one risk factor (low income). While the authors suggested that the groups were equal because parent education and fathers in the home were the same, among other variables, upon further examination there were several significant risk factors present in the treatment group at a rate much higher than the control group. These included limited English and language deficiency, unemployment, no support system, substance abuse, family delinquency and incarceration (Maloffeva et al., 2007). While the control group scored worse on several other risk factors as well, it is possible that unemployment and language may have been determining differences. By seventh grade the Pre-K group had a significantly higher rate of special education placement compared to the no Pre-K group. This is hard to explain without unequal groups, unless we conclude that Pre-K quality was so low as to correlate with needing more special education (i.e. harmful) at some point in the future, which is nearly impossible given other positive and significant findings in the studies related to retention, social relations and fourth grade literacy and math results (Xiang & Schweinhart, 2002). An alternative hypothesis for Michigan is that the yearly compounding nature of having two risk factors in the home rather than one at all times, and the nature of those risks, eventually separated the groups with results not visible until middle school. Clearly, group equivalence will remain a concern in quasi-experimental studies that lack robust matching methods.

Strong designs controlled for child-level and school-level covariates in their regression models. Furthermore, England stood out for controlling for home environment on a wide range of factors (Sylva et al., 2011) becoming the first to measure home learning environment in relation to preschool quality. Ten studies used more advanced

regression methods such as ordinary least squares (OLS) and hierarchical linear modeling (HLM). Two studies used ANOVA or ANCOVA which provide less flexibility in the analysis of the data (Valenti & Tracey, 2009; Xiang & Schweinhart, 2002).

Results of Pre-K on later child outcomes

All of the international studies reported significant results of Pre-K on some academic measure of later child achievement during the final time period of data collection. Of the U.S. studies, seven reported some academically significant finding. However, results were rarely significant across the board for all measures. Normally, one outcome would persist in significance (i.e. lower retention) while achievement tests were more likely to fade. The most persistent result overall was that of England, finding significant results through age 11. The most persistent result in the United States was Michigan's Great Start Readiness Program which found significant results in fourth grade, but these results did not persist into seventh grade, and there are concerns about group equivalence in the study. Taken together, 12 of 16 studies (75%) reported at least one statistically significant result for at least one academic outcome (math, language or literacy) in the final year of testing. Most longitudinal studies find some significant outcome of state-funded Pre-K attendance on later grade school academic achievement beyond kindergarten readiness. See table 2.1 and 2.2 for a summary of this literature.

Table 2.1

List of longitudinal evaluations of state-funded Pre-K programs reporting results past school readiness (2000 – 2012)

Study / Year	Location / Program	Design / Statistics	Duration	Outcome Domains
Sylva 2012	England *EPPE*	Mixed, Multi-level modeling	Age 11	Math English Social and Emotional
Sylva 2004	England, *EPPE*	Multi-level modeling	End 1st	Reading Math
Aguilar 2012	Uruguay *Universal Pre-K*	OLS Bivariate Probit	K 5th	National Test (Domains not stated)
Huang 2011	Virginia *VPI*	Quasi 2L HLR (5)	End 1st	Literacy K retention
Valenti 2009	Northeast *School District*	ANOVA Levene's Test	Beg. 1st Mid 1st	Literacy (DRA2 reading level)
Berlinski 2009	Argentina *Universal Pre-K*	Modeling, OLS regression	3rd	Spanish Math Self-Control
Frede 2009	New Jersey *Abbot*	Quasi, Selection Bias & Attrition.	End 1st, End 2nd	Literacy Oral Lang. Reading Comprehension Math K-1 retention
Frede 2007	New Jersey *Abbot*	Quasi, same as Frede '09	Beg. K End K	Oral lang. Early literacy Math
Hustedt 2008	Arkansas *(ABC)*	Quasi	End 1st	Language Literacy Math
Peisner-Feinburg 2008	North Carolina *More at Four*	Various Regression.	End K	Language Literacy Math Behavior General Knowledge
Waldfogel 2008	Int'l *7 OECD countries including US*	OLS (Effect of money spent in Pre-K on 4th grade scores.)	4th	Math Science
Magnuson 2007	National ECLS-K Data	Multivariate OLS Regression	End 3rd	English, Math
Magnuson 2007	National ECLS-K Data	Propensity score matching; OLS	Beg. 1st	Literacy Math

Table 2.1, Continued

Study / Year	Location / Program	Design / Statistics	Duration	Outcome Domains
Maloffeva 2007	Michigan *Great Start Readiness Program (GSRP)*	HLM H2LM	6th 7th 8th	Reading Writing (7th grade only) Retention Special Edu. Attendance Course selection Services
Xiang 2002 & Xiang 2000	Michigan *GSRP*	ANCOVA Regression	2nd 4th	Literacy Math Social Retention
2L HLR: Two-level hierarchical linear modeling.				

Table 2.2

List of longitudinal evaluations of state-funded Pre-K programs reporting results past school readiness (2000 – 2012)

Study / Year	Curricula Named	Preschool Quality Measured (3)	Later School Quality Measured	Fading Effects Analyzed (2)	Significant Results (1)
Sylva 2012	No	Yes	No	No	Yes
Sylva 2004	Yes	Yes	No	Yes	Yes
Aguilar 2012	No	No	Yes (4)	Yes Weaker but positive.	Yes Not described
Huang 2011	No	No	No	Yes	No
Valenti 2009	No	No	No	No (Sleeper effects)	Yes
Berlinski 2009	No. Purpose mentioned	No	No	No	Yes
Frede 2009	No	Various standards of quality are met. No empirical data.	No	Possible catch-up effects. No empirical analysis. (6)	Meaningful to significant.
Frede 2007	No	Same as Frede 2009	No	No, stated to persist.	Yes
Hustedt 2008	No	Yes	No	No	Yes

Table 2.2, Continued

Study / Year	Curricula Named	Preschool Quality Measured (3)	Later School Quality Measured	Fading Effects Analyzed (2)	Significant Results (1)
Peisner-Feinburg 2008	No. Sites self-report using an official curricula	Yes	Yes	Yes, but short duration limits analysis / results	Yes
Waldfogel 2008	No	No	No	No. Says that low kids gain more.	A $100.00 increase in funding adds about .05 SDs in math / sci.
Magnuson 2007	No	No	Yes (7)	Yes	Yes
Magnuson 2007	No	No	No	Yes. Persist more for at-risk.	No
Maloffeva 2007	No	Yes	No	No	No (tests) Mixed on others
Xiang 2002 & Xiang 2000	No	Yes	No	No	Yes (4th grade)

Only studies meeting the criteria for inclusion are reported here. Studies from the same program are grouped together regardless of year.

Notes:

(1) Pre-K was found to have a significant effect on child academic outcomes at the final time period studied. Not all outcome domains may have been significant during at or along the time periods studied. See the particular study for exact findings.

(2) To analyze fading effects the study needed to report outcome data at least twice. Studies reporting an outcome at a later age only once would not be able to determine fading effects, lacking an initial comparison of longitudinal impact.

(3) This indicates that preschool quality was measured on at least a subset of the sample population. Please refer to the exact study for measures and methods of preschool quality. Observations using ECERS-R, among others, were typical.

(4) One variable predicting school quality (school average rate of retention across the time period of 1999 – 2004) had a stronger effect on performance at year 6 (5th grade).

(5) Teacher survey of the number of minutes spent in reading instruction greater than 90 minutes per day and classroom size of 20 or less students indicated later school quality. Other school effects accounted for in the statistical model.

(6) Catch-up effects are hypothesized because of the high focus in the district on literacy and the average test results of all children. This suggests that later school quality is good.

Conclusion

Studies made very little effort to analyze Pre-K curriculum usage in relation to later child achievement. In the United States, this is partly due to the flexibility that local communities are given to choose their own curriculum. However, it makes less and less sense to support such flexibility as evidence mounts that most preschool curricula are rather ineffective. When sites do choose from a list of approved options, the amalgamation of different curricula being used across the same region is difficult to separate for any useful analysis on later achievement (Henry, Gordon, Henderson, & Ponder, 2003). Perhaps researchers feel that the measurement of Pre-K quality can substitute for the mention of the particular curriculum, but such an omission does a disservice to readers examining evaluative reports for guidance on how to improve or implement their own programs.

The research in England (Sylva, 2011) suggests that medium quality Pre-K is a necessary threshold to reach in order to see academic test score results significantly improve compared to a home group at age 11 (fifth grade) while high-quality Pre-K is a necessary threshold to reach to see all measures of social and emotional development significantly outperform all other student groups. High-quality Pre-K also maintained a statistically significant advantage over medium-quality Pre-K, at age 11, on math and English test scores suggesting that it is well worth the effort to improve Pre-K programs to the high mark. Her results suggest that low-quality programs provide little value from a child development perspective (as opposed to a parental child care perspective).

Further credibility is lent to the theory that the effect of Pre-K is influenced by later school quality (Aguilar & Tansini, 2011; Magnuson et al., 2007a). (This did not include the effect of instructional quality, which was absent from these studies.) After controlling for school quality, Magnuson et al. (2007a) found significant results of Pre-K attendance on children's third grade academic test results. This supports the work of educators to align quality across the PK-3 continuum to better advance child academic trajectories. Findings support the notion that child growth trajectories can be changed by maintaining quality, year over year, to third grade.

Limitations of the current review

This review is limited by potentially missing evaluations not found during the search process. However, the number of studies meeting the criteria is clearly small. It is unlikely that studies of a significantly higher design quality would be found with results that fundamentally alter the conclusions. Another limitation is the narrow scope of this review, as set by the conceptual framework which focused analysis on a set of criteria. Other variables may be of interest to readers, and other researchers may prefer to use different criteria to analyze these evaluations. Nevertheless, the criteria used here to evaluate the studies is grounded in evidence and succeeded in revealing gaps related to duration, preschool quality control, grade school quality control, curriculum evaluation and design questions.

A general caution regarding longitudinal research is the fact that programs change significantly as they evolve from year to year, and personnel change. A later cohort of

children may experience significantly higher (or lower) program quality compared to previous cohorts, rendering longitudinal findings obsolete by the time they are published. As such, longitudinal evaluations of the nature described here provide evidence of the effect of Pre-K several years ago. However, to the extent that studies reported empirical data on the quality of the Pre-K programs we can infer that similar results will be repeated, so long as current quality is of equal or higher value.

Implications for research and practice

There is hardly any evidence that current longitudinal evaluations are determining if Pre-K programs use evidence-based curricula. Would the use of an evidence-based curriculum be related to longer-term child outcomes? To what extent does local choice over curricula correlate to later child outcomes? Pre-K programs need to ensure they are using an evidence-based curriculum, such as those reviewed by the What Works Clearinghouse, and future evaluators of Pre-K need to explicitly examine curricula in relation to their findings.

This review found some evidence to support the sustained effect of preschool when later grade school quality is higher. The implication for practice, policy and research is to align a high quality Pre-K program with a higher quality grade schooling system, an obvious goal.

Meanwhile, future studies need to statistically split, and separately report, the effect of Pre-K on later child outcomes by domains of mediocre versus effective elementary schools. Pre-K should not be held accountable for longer-term results if later

school quality is found to be low. The literature will be strengthened as more researchers stratify and report results according to Pre-K quality profiles, taking note of Sylva (2012). No study has yet used quality rating and improvement scores (QRIS) as the basis for creating a range of Pre-K quality profiles from which to test child outcomes in later grade school. This would require evidence of validity and reliability of QRIS as an instrument. The only study to create profiles of quality and then compare each profile to later outcomes was the England study (Sylva, 2012).

Policy makers and agencies need to consider the fact that very few longitudinal studies can be completed at all in an environment of disjointed data systems, disconnected institutions of early childhood, and lack of common reporting of achievement statistics across the PK-3 grades (Hernandez, 2012). The small number of evaluations found by this review confirms that a more friendly evaluative environment needs to be created through integrated data systems, and cooperation between researchers, policy-makers and practitioners on what achievement variables need to be included from year to year to make longitudinal research more possible and to better tie the PK-3 experience together in the interest of children. Recognizing these difficulties, the current review found that few evaluations extend the search for the sustained effect of Pre-K beyond first grade. Future studies should make every effort to cover the entire PK-3 range.

Summary

As perhaps the first literature review on the longitudinal effect of state-funded public prekindergarten beyond school readiness, this chapter has added to the knowledge base in a number of areas. Seventy-five percent of all Pre-K evaluations show some positive effect on child academic outcomes, strengthening the knowledge base surrounding the ability of Pre-K to make a measurable impact on child outcomes beyond kindergarten and as high as fifth or sixth grade. A reasonable expectation for Pre-K sustained effects is first grade with a goal of third grade, but only if later school quality is taken into account and only if Pre-K quality is measured as good rather than average. The literature already establishes that average Pre-K has an effect on school readiness. This review extends that understanding with implications to the entire PK-3 continuum. The need to control for grade school quality, measure and control for Pre-K quality, use proper matching methods in designs, extend evaluations into first grade and beyond, and include more data points along that continuum, are all further areas of contribution to the field of longitudinal research in early childhood education.

Very little work has been done by researchers to control for home environment variables, evaluate curriculum within Pre-K effect studies, include observed measures (as opposed to teacher report measures) of social and emotional outcomes, and use "true value" tests of latent ability rather than standardized tests when studying the longitudinal effect of Pre-K. School district level studies are hardly represented in the literature. Most Pre-K programs have yet to be evaluated longitudinally to test for results beyond school readiness, as shown in the small number of studies that could be found by this review.

While it is expected that preschool would have an impact on school readiness as a direct consequence of operations, the ability of Pre-K to extend the duration of a child's learning for several more years remains an open question in need of further exploration. The current study will be the first at the district level to do all of the following: use advanced propensity score matching to approximate group equivalence, include balanced outcome measures between cognitive and affective domains, control for later school quality in relation to particular findings, and describe the specific Pre-K curriculum that has been used.

CHAPTER 3

METHODOLOGY

Introduction

Longitudinal research helps ensure that quality is raised and sustained over consecutive early childhood years (Nelson, Westhues, & MacLeod, 2003) allowing the changes in child achievement to be monitored for targeted improvement from grade to grade. However, very little research on the effect of state-funded Pre-K in the United States has investigated impacts beyond school readiness. The scant evidence for Pre-K sustained effects naturally emphasizes early literacy, but just 30% of such studies measured children's later social or emotional development.

Research Questions

1. Does attendance in New Day Pre-K make a sustained difference in children's literacy achievement in first grade?

2. Does attendance in New Day Pre-K make a sustained difference in children's social and emotional adjustment in first grade?

Purpose of the Study

The purpose of the study is to observe the effect of Pre-K attendance on children's later cognitive and affective development in first grade. This study also aims to address a concern in the literature on the later effect of Pre-K, in which later school quality is rarely taken into account in study designs (chapter 2). The current study exerts a level of control over later school quality in the design. The study also reports the specific Pre-K curricula used in the district, knowing that Pre-K curricula have varying degrees of effectiveness (What Works Clearinghouse, 2013) and that virtually no other study on the later effect of Pre-K has reported the specific curricula being used.

This study seeks to contribute to the understanding of the effect of Pre-K attendance on children's later literacy and social/emotional development. While both of these outcomes are crucial to children's overall development, later behavior and social adjustment outcomes as a result of typical Pre-K attendance is rarely studied and remains a key question in the field.

Research Design

A quantitative ex post facto (causal comparative) study design was used. As a retrospective cohort study, archived data were used to observe changes between the control and treatment groups based on an exogenous independent variable: Pre-K attendance. It is retrospective in nature because all events associated with the data occurred in the past, and all analysis occured within a single shortened time period.

Rationale for Design

A quasi-experimental design was chosen for several reasons. First, it is not possible to randomly assign children to the treatment of preschool because this is infeasible. Secondly, data existed on who attended New Day Pre-K and who did not in the city. Thirdly, a quasi-experimental design with propensity score matching to establish baseline equivalence between the treatment and control group was chosen because it addresses selection bias and reduces it to the minimum. Usually, a pre-test and post-test would be used to compare the groups before and after treatment, but there is not a pre-test available in this case, and it was not possible to test all children, including control children, at age four (at baseline). Therefore, this study was a post-test only comparison with equating, where the samples were matched to make sure they were equivalent at baseline using available demographic data.

Approximating Group Equivalence with Propensity Score Matching

Grouping (control vs. treatment) was completed using a matching method. Children who attended Pre-K were matched with similar children who did not attend any type of formal or institutional preschool, such as Pre-K, Head Start or private providers. The method of matching to approximate group equivalence is Propensity Score Matching (PSM) which assigns a single score (the propensity score) to every participant representing their likelihood of receiving the treatment. PSM allows group assignment to

be completed by controlling for a large number covariates, rather than matching on one particular variable. This reduces the error introduced by selection bias.

If group sample sizes were relatively equal, then nearest neighbor matching (one-to-one) could have been employed. However, the group sizes in this study were widely different from one another, so optimal matching using the MatchIt package (Ho, Imai, King & Stuart, 2011) with the R analysis software on a one-to-many fixed ratio was performed, where each control unit was matched to several treatment units (1:5 ratio). This still preserved the treatment sample by matching each control unit to five treatment units. Covariates were matched on:

1. Age
2. Black
3. Asian
4. Mixed
5. Black and White
6. Hispanic and White
7. Male
8. Speech/language disability
9. Learning disability
10. Eligible for free lunch
11. Eligible for reduced lunch

If a student was coded as a 0 for all races listed, then he/she was White and was matched accordingly. This process was used for each demographic category. For example, Male would was coded as 1 or 0. If a student was 0 for Male, the gender was female, and so on for each covariate in the matching process.

Variables

<u>Independent</u>

The independent variables (treatment group) were: 1) attended the public

preschool (New Day Pre-K Program) and 2) did not attend New Day Pre-K or any other

type of formal preschool, such as Head Start or private preschool.

<u>Dependent Variables</u>

One dependent variable in the study was student achievement in literacy (reading

ability by text level, spelling, and a composite measure). Another dependent variable was

behavior, as a proxy for social and emotional development, classified as teacher reported

citizenship skills taken from report cards. This was used as an indicator of social

adjustment in school.

Behavior Composite Index

A composite behavior score was calculated for each child. This behavior sum

score was then used as a scaled index variable to conduct outcome analysis on the effect

of New Day Pre-K attendance on social adjustment in first grade. For the behavior

composite variable a score of zero is perfect, meaning zero N's were given on the child's

report citizenship report card. Each N ("needs improvement") was assigned a score of 1,

Therefore, any results found using this behavior index were not confounded by income or race, for example.

Variables

A number of variables were available for use in statistical analysis. From this list, only those covariates that were considered to be continuous or stable from Pre-K to first grade were used to conduct propensity score matching marked with an asterisk (*), below. Those without an asterisk are available for subgroup analysis and interaction effect analysis for future research.

1. Age (as raw score: e.g. 6 or 7) *
2. Race (race and ethnicity codes) *
3. Gender *
4. Special education status (disability) *
5. Socio-economic status (income: as free, reduced or full price lunch) *
6. Retention (and grade level retained, such as K or 1, or multiple years of preschool)
7. Mobility proxy / transfer (number of years consecutively enrolled in each previous school and name of each previous school)
8. Zip code (as proxy for neighborhood environment)
9. School name (for each consecutive grade)
10. School quality by state test results in math and reading
11. School Title I status
12. Number of years in Pre-K (one or two years of preschool, as some children start at age 3)

School Quality Index and Profiles

Elementary schools were categorized as falling into one of five quality profiles based on the combined score of their math and reading standardized state test results

reported from third, fourth and fifth graders. This created an estimate of academic success in each school, allowing for schools to be compared one to another using a composite test score. A perfect reading or math test score for a school was awarded 100 points, representing a 100% pass rate on that test by all students in the school. The highest score a school could receive, therefore, was 200 points: 100 for math and 100 for reading, which would be a perfect passing rate on both subject tests combined for all students in the third, fourth or fifth grade in that school for the 2011 – 2012 academic year. After calculating a composite test score for each school, schools can be compared across the state one to another, and ranked in order from first to last on test results. This represents only one generalization of quality and does not include information from site visits, or direct observation using observational measures, and there is no guarantee that kindergarten or first grade quality in the same school was equally excellent, good, fair, low or very low. Without individually scoring kindergarten school quality in every school in the district, the alternative is to use the standardized test results as a proxy or estimate for general or overall academic effectiveness in a school. This also does not imply that the social and emotional climate or quality of the school is equally good.

The school quality sum score was then used to classify each school as falling into one of five quality profiles (Table 3.2). This school quality index is on a 1 – 5 point scale, where 5 is excellent and 1 is very low. An ascending score index, such as this, permits multiple regression analysis on later child outcomes by school quality to be performed in follow-up research. This study generated results by quality profile using the Split File function in SPSS, separating the data into the different quality groups, and then compared the means of the control and treatment groups within each quality subgroup to determine

if there was a significant difference on later child outcomes by school quality. An example of how schools were scored and categorized using the quality index is shown in Table 3.2.

Table 3.2

School Quality Score and Index (Sample)

School	Reading	Math	Total Score	Quality Profile Index: 5 = Excellent 4 = Good 3 = Fair 2 = Low 1 = Very Low
1	83	63	147	3
2	90	74	164	4
3	72	36	108	1

Population and Sample

The population was all 2,221 first grade students in the district during the 2012 - 2013 academic year. The population was of mixed race in a primarily urban setting.

The original sample consisted of 1598 students (treatment, n=1269; control, n=329). Other students in the population were excluded from the sample (n=623) because they received a different form of preschool (Head Start, private provider, etc.) or had no preschool information reported.

Sample Reduction Process

The treatment and control samples of 1269 and 329 were further reduced to 1206 and 284, respectively, due to missing data. Then, this control group of 284 children was reduced again to 176 to exclude children who transferred into New Day Schools from outside the city. Excluding children who transferred from outside the city was necessary because of the likelihood that children from another city may have been incorrectly surveyed as having no formal preschool experience. The study sample at this point contained 1382 cases (T=1206; C=176).

At this point, any child who had all of the outcome scores missing was deleted because such children would not be part of the analysis if they were missing all outcome data. This could have occurred because a student may have missed an assessment, or transferred between schools in the middle of first grade or because of a data entry error at the division level. By deleting cases with significant missing outcome data, the sample was reduced again, from 1382 to 1373 cases. Those cases with a few missing outcome values were still included, so long as they were not missing all outcome values. This reduction of 9 students for missing outcome data occurred in the treatment population, so the final sample before matching was: T = 1197; C = 176. Each control case was then matched to five treatment cases using the "optimal" propensity score matching technique, which results in a final analysis sample of 176 control cases and 880 treatment cases, or 1056 children overall.

Data Collection

Demographic Information

Demographic information was collected from the Student Information System electronic database provided by the school district. These variables were itemized above.

Achievement Data

Academic achievement data were collected from the school district's office of research. Archive data on student reading achievement existed from two measures: *Phonological Awareness Literacy Screening* (PALS) as text levels, sight words and spelling scores and the *Developmental Reading Assessment* (DRA2) as independent reading level.

Assessment data of math, writing, science or social studies was unavailable from the district. A later prospective study (to 3rd and 5th grade) could include these domains using state testing results that would become available when the current first grade cohort (2012 – 2013) reaches the end of third grade by the spring of 2015.

Affective Data

Data related to children's social and emotional development was collected from report cards based on teacher report of children's use of citizenship (cooperation,

response to authority, following rules, etc). This data was collected from the electronic records of archived report cards in the district.

Instruments & Materials

<u>Reading Achievement</u>

Phonological Awareness Literacy Screening (PALS)

The PALS is a criterion-referenced assessment given twice a year (fall, spring) to each child in a one-on-one conference with the teacher. It identifies if children are at-risk of reading difficulty or not at risk of reading difficulty according to grade level expectations. It is administered, with various subtests suitable to the grade level of the child, from preschool (PALS-PreK) to kindergarten (PALS-K) and through grade three (PALS 1 – 3). PALS assessments are used by every school in the state and are reported to the state department of education. The PALS has advantages over the DRA2 for young children because it includes assessment of phonological awareness, rhyme, and concept of word for emergent readers.

Developmental Reading Assessment, 2nd Edition (DRA2)

The DRA2 measures student's ability to read fiction and nonfiction texts. The test includes sub-scores for accuracy, fluency and comprehension on a passage of text that

has been read by the child to confirm a particular independent reading level. Reading levels begin with 1, 2, 3, and 4, and then rise by twos, as 6, 8, 10 into the 20's, and then rise by fours beyond level 30: as 34, 38, etc.

The DRA2 is a commonly used teacher measure for student reading performance and progress. School districts provide training to teachers in the use of this measure. It is administered in a one-on-one conference between the teacher and child. Children read a leveled text while the teacher records errors and notations on an observation form with the same text. The student's errors are divided by the number of words read to determine a rate of accuracy, with no errors being 100% accuracy. Reading rate (fluency) is measured by timing the speed at which a student finishes a passage, for text levels 14 and above. Comprehension is measured by student oral responses to questions (below level 28) or through student written responses for text levels 28 and above. A student passes a particular text level, or is considered an independent reader at a particular text level, when their accuracy, fluency and comprehension scores all exceed a stated benchmark score for each of those three constructs. Student reading achievement is indicated by their ability to independently read and comprehend each proceeding text level in the DRA2 continuum, which includes texts for K – 8.

Inferring DRA2 Scores Using Report Cards: Data Imputation as Needed

The school district in the study assigns a reading performance grade for all students on quarterly report cards based on the most recent PALS and DRA2 text level results. Specific PALS and DRA2 text levels are required to receive a below, on grade

level or above grade level mark on report cards. A students' DRA2 score can be inferred based on the report card results by using the district's correlation of grade-to-expected DRA2 levels. For example, a first grade student is expected to be reading a DRA2 level 14 by the third marking period of the year. Whether the student is reading a DRA2 14 can be determined by the report card data indicating if the child is "on grade level." This allows DRA2 reading levels to be inferred using report cards as a potential method of imputation if reading data are missing.

Affective Outcomes

Behavioral outcomes will be classified as citizenship skills based on teacher report taken from report cards. Citizenship grades will serve as a proxy for social and emotional development.

Validity and Reliability

PALS (Reading)

The PALS assessment has a high classification accuracy with an area under the curve (AUC) of .91, meaning it accurately diagnosis's children as either at-risk or not at risk of reading difficulty (meeting or not meeting minimum literacy benchmark). Cronbach alphas for the PALS 1 – 3 range between .79 and .93. Interrater reliability has been tested between .98 to .99 (Huang, Invernizzi, & Drake, 2011).

DRA2 (Reading)

Reliability analyses performed for the DRA2 include internal consistency

reliability (.50 to .80 reliabilities between fluency and comprehension), passage

equivalency (MANOVA used to show no significant differences), test-retest reliability

(correlation coefficients above .90) and interrater (66 – 72% agreement) and expert rater

reliabilities (McCarthy & Christ, 2010). The DRA2 was tested for validity using

criterion-related validity (no significant difference with other test: with .60 - .70

correlations), construct validity (low correlation at .41 across subtests) and predictive

validity (teacher ratings with DRA2 scores: coefficient .60 to .63) (McCarthy & Christ,

2010).

Affective Outcomes

Subjectivity in report card grading to communicate achievement data is a well-

documented concern (Linn & Gronlund, 2000). Using report card data to assess

children's social and emotional development through citizenship grades on specific skills

(following the rules, responding appropriately to authority, respecting the rights of peers,

cooperating with peers, etc) is limited to the subjective view of the teacher and faces risks

related to reliability and validity. However, the children in this district do not receive a

single grade for citizenship. Rather, they receive seven separate grades in citizenship, one

for each skill area. The possibility that a child with poor behavior, and who lacks

prosocial skills in a noticeable way, would still earn all "S's" on all seven citizenship

skills is highly improbable. Subjective teachers will most likely assign at least one "N" to one of the skill areas when the child has behavioral challenges. Although a teacher's report of social adjustment on any one particular skill is subjective, the likelihood that a poorly adjusted child would earn "perfect" citizenship marks is very low.

In addition to the improbable chance that a poorly behaving student would earn 7 "S's" in citizenship, teachers may subjectively assign all S's to a child simply because he/she is generally a "good kid." However, the purpose of this study is not to measure exactly what prosocial skills a child has or to what degree he/she possesses them. This study determines if there is a statistical difference in the number of S's, N's or U's that children in first grade receive on their citizenship report card as a function of attending or not attending New Day Pre-K.

Data Preparation and Coding Procedures

The different levels of demographic variables, such as SES, race/ethnicity, gender and disability status, were coded into dummy variables to allow for PSM matching or post-matching analysis with them to take place. Impact analysis was performed using all the cases with valid data on a given outcome variable. The analysis sample varies slightly depending on which outcome variable was included in the tests because not every child had complete outcome data across all values. Children with a few missing outcome values were still included in the sample. It was necessary not to delete cases entirely due to a few missing outcome variables in order to preserve sufficient power to detect the effect. The analytic samples are as representative of the population as possible.

Data Analysis

Inferential statistics were used for data analysis following matching. Because of significantly balanced groups, a series of independent-samples t tests were used to test between group differences ($\alpha = .05$). In this case there were two groups being measured on a single continuous outcome variable. T tests of significance were then converted to effect size scores to determine the magnitude of the difference between the groups. The effect size conversion of the p-value scores gave meaning to the results because it calculated how great of an effect Pre-K attendance had on each particular dependent variable.

If there was no significant difference on a particular outcome variable, subgroup analysis by school quality in kindergarten was performed to determine if there was an interaction effect between the outcome variable and later school quality.

Limitations

An ex post facto design does not directly manipulate treatment (is not experimental). The study lacked random assignment, introducing unobserved selection bias. However, propensity score matching on a wide range of variables, with a large sample size, approximates group equivalence in the absence of random assignment (Stuart, 2010). As a causal comparative study, the analysis still permitted a degree of causal inference.

Certain confounding variables were unknown, such as parent's highest level of education (although this is related to SES) and home environment. Parents who decided to enroll their child in Pre-K may have been fundamentally different from parents who did not enroll their child in Pre-K in ways not accounted for by the covariates available to the researcher. For example, the control group could have been significantly more disadvantaged than the treatment group in ways that race, SES, or disability would not reveal. However, all available baseline covariates were used in the matching process and the standardized difference between the two groups was reduced to five percent which is significantly below the 10% threshold.

Results may not generalize to other cities, states or regions in the United States because data were taken from one urban school district. However, many other cities have similar characteristics with regards to diversity, income and urban concentration, which gives significance to this study as a model for other districts considering expansion and evaluation of Pre-K programs. To strengthen generalizability a conservative effect size calculation was used.

Longitudinal research faces the limitation of program change, which hinders the ability to replicate findings. However, this cohort of children were enrolled in Pre-K in 2010/2011 when there was a significant focus on literacy, which remains the case today.

Using report card grades on seven citizenship skills to measure behavior was a limitation due to subjective teacher-report (Linn & Gronlund, 2000) and the absence of direct observation or testing conducted by the researcher. However, the composite behavior index that was created for this study totals children's entire report card skill scores to determine if there was a significant difference in the total number of various

citizenship grades between the control and treatment groups. Finally, the reading assessments were administered by teachers introducing the possibility of teacher subjectivity. However, the DRA2 and PALS 1 – 3 tests have acceptable reliabilities.

Protection of Participant Rights

All of the electronic records and dataset files were stripped of personally identifying information before being received by the researchers. Anonymity was also protected by reporting results in the aggregate. The files contained a common code as an identifier to merge various datasets, which was the student's unique identification number. No files containing student's names or addresses linked to the identification number were available to the researcher. The study used all archived data, and posed no risk to participants.

Summary

As a retrospective study on the effect of Pre-K attendance by first grade, this research makes a significant contribution to the field of early childhood education regarding PK – 3 alignment and quality efforts, as well as Pre-K evaluation efforts and issues related to later school quality and Pre-K attendance. The study design included a reasonable matching method to improve balance between control and treatment groups. A large number of covariates are controlled in the design, including key variables such as income, race, gender, age and disability.

By including a proxy for student social and emotional development as a dependent variable, in addition to literacy, the study contributes to a more balanced view of child development rather than evaluating Pre-K attendance on later academic achievement alone. The study may also be the first in the United States to judge Pre-K sustained effects by different school quality profiles in the case of behavioral outcomes.

The study is also well-positioned for follow-up research on this cohort of children (the 2010/2011 preschool class) as they progress into second, third, fourth and fifth grade. A prospective study of the cohort in 3rd and 5th grade is possible.

This particular public school district has nearly universal Pre-K access, which remains rare in the United States outside of Oklahoma, Florida and Georgia (Barnett et al., 2011). As a case study in large-scale, district provided Pre-K, this research is potentially valuable to numerous districts across the country, and particularly valuable to policy-makers who are considering how to expand or improve their state Pre-K programs.

The expected trend to expand Pre-K access in the near future, as the economy stabilizes, makes this study timely and useful. Finally, the results of the study will help inform the participating school district about the ability of their Pre-K program to have longer-term impacts, which is a question that they do want to answer. Furthermore, this study can be replicated by the district in future years to examine the effect of any new Pre-K curricula or standards that may be adopted over time. This study is relevant to the local district and to national questions about Pre-K access and PK – 3 development.

CHAPTER 4

DATA ANALYSIS & FINDINGS

Introduction

The purpose of this study was to investigate the possible effects of attendance in the New Day Pre-Kindergarten program on first grade literacy and behavior outcomes in an urban school division in the southeastern United States. Chapter 3 explained how data were coded and prepared for analysis. This chapter begins with the results of the propensity score matching procedure demonstrating increased group balance to a sufficient level necessary for analysis to take place. The findings are then presented based on the research questions and effect sizes are reported to determine the magnitude of the effect for significant outcomes.

Matching Procedure and Approximation of Group Balance

Optimal matching with a set control-to-treatment ratio of 1:5 was used. The control group was previously reduced to 176 cases to exclude children who had transferred from a different city, as mentioned in chapter 3. After the fixed-ratio matching, 176 control cases were matched to 880 treatment cases (176*5=880; sample size of 1056). A total of 317 treatment cases remained unmatched and were computationally excluded.

The pre-match table (below) shows that the absolute standardized difference between the treatment and control groups on key matching variables was initially 12.12, with a control group of 176 cases and a treatment group of 1197 cases. Any average standardized difference score above 10 indicates that the two groups are significantly different (unbalanced), and that matching should be conducted to reduce the absolute standardized difference to sufficiently balance the two groups on the covariates included in the matching model.

Propensity score matching using the optimal matching method on a 1:5 fixed ratio was then conducted. First, the statistical analysis software, R, was installed. Then, the MatchIt package used for optimal matching, for use with R, was installed. The syntax command codes to perform the matching in R, calling the MatchIt package, were customized based on the variables applicable to this study and the commands were then run. Before running the syntax commands, a .csv data file was imported into R from a cleaned excel data sheet containing all matching variables. After R ran the matching commands, an output .csv file was generated and converted to an excel file. The post-matched data was then saved as an SPSS file and merged with the outcome datasets to create the final matched SPSS file for analysis.

Results of propensity score matching show that the average standardized difference between the two groups was reduced from the pre-match level of 12.12 to the post-match average of 5.1, which is sufficiently balanced to conduct analyses using an independent-samples *t*-test (see table below). Only one covariate remains above the 10% threshold (Hispanic and White). If one or two covariates have an average standardized difference above 10% after matching (such as the Hispanic and White race variable at

14.94) the two groups are still considered to be sufficiently balanced on the whole, and analysis can proceed so long as the average standardized difference is below 10% (Rosenbaum & Rubin, 1983). By reducing the average standardized difference to below 10%, as a result of propensity score matching, balance between groups was increased significantly to allow outcome analysis to proceed using a t test. In the absence of this successful optimal matching on the propensity score, the study would have had to control the covariates through post-match regression to conduct outcome analysis.

Table 4.1

Pre / Post Match Balance Results

Pre-Match	Control (N=176)		Treatment (N= 1197)			
Covariates	**Mean_T**	**SD_T**	**Mean_C**	**SD_C**	**Standardized Difference**	**Absolute SD**
Age	6.55	0.57	6.52	0.53	5.55	5.55
Black	0.41	0.49	0.57	0.50	-31.50	31.50
Asian	0.05	0.21	0.03	0.17	9.03	9.03
Mixed	0.09	0.29	0.10	0.30	-2.05	2.05
Black_white	0.05	0.22	0.04	0.19	6.14	6.14
Hisp_white	0.10	0.30	0.05	0.21	21.58	21.58
Male	0.57	0.50	0.51	0.50	12.57	12.57
Speech	0.02	0.15	0.03	0.18	-6.47	6.47
Learning	0.01	0.11	0.01	0.11	-0.31	0.31
Free	0.69	0.46	0.62	0.49	13.88	13.88
Reduced	0.03	0.18	0.09	0.29	-24.21	24.21
					Average	**12.12**

Post-Match	Control (N= 176)		Treatment (N= 880)			
Covariates	**Mean_T**	**SD_T**	**Mean_C**	**SD_C**	**Standardized Difference**	**Absolute SD**
Age	6.55	0.57	6.54	0.53	2.67	2.67
Black	0.41	0.49	0.46	0.50	-9.61	9.61
Asian	0.05	0.21	0.04	0.19	4.58	4.58
Mixed	0.09	0.29	0.11	0.31	-6.78	6.78
Black_white	0.05	0.22	0.04	0.20	3.75	3.75
Hisp_white	0.10	0.30	0.06	0.24	14.94	14.94
Male	0.57	0.50	0.59	0.49	-2.76	2.76
Speech	0.02	0.15	0.02	0.15	0.77	0.77
Learning	0.01	0.11	0.01	0.09	2.25	2.25
Free	0.69	0.46	0.66	0.48	6.77	6.77
Reduced	0.03	0.18	0.03	0.18	1.27	1.27
					Average	**5.10**

*Standardized difference =
$$\frac{100(\bar{x}_{treated} - \bar{x}_{control})}{\sqrt{\frac{s^2_{treated} + s^2_{control}}{2}}}$$

Matched Sets

The PSM procedure produced matched sets of six cases per set (1 control, 5

treatment). After matching, 176 matched sets were created, inclusive of the complete and

final sample (176*6 = 1056). The following example shows how student ID numbers

were grouped into matched sets using the propensity score, where subclass represents the

number assigned to each matched set. Treatment (T) = 1; Control (C) = 0. This example

illustrates that matched sets were formed using the propensity score, and that the resulting

dataset is ready for analysis.

Table 4.2

Example of Matched Sets Following Propensity Score Procedure

ID	T or C	Propensity Score (distance)	Matched Sets (subclass)
17C8Q3	1	0.1145	172
17G6A6	1	0.1145	172
18K7L9	1	0.1145	172
17P9E1	1	0.1145	172
18M7Z7	1	0.1145	172
30U4H5	0	0.1145	172
17B5O6	1	0.131	171
17A2Y0	1	0.131	171
18N7E3	1	0.1259	171
17V3W1	1	0.124	171
18R4M8	1	0.131	171
30T3J9	0	0.131	171

Reading Outcomes

Does attendance in New Day Pre-K make a sustained difference in children's later literacy achievement in first grade? Descriptive and inferential statistics are presented for reading outcomes based on six measures of literacy achievement in the first grade.

Descriptive reading outcome statistics show that the treatment group, listed as "1", has a higher average reading score compared to the control group for every literacy test reported, in both the fall and the middle of the year (Table 4.3). Results are reported in narrative form in the following sections.

Table 4.3

Reading Outcomes: Descriptive Statistics

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Fall PALS Text Level	0	172	3.07	1.869	.142	2.79	3.35	1	10
	1	865	3.86	2.102	.071	3.72	4.00	1	10
	Total	1037	3.73	2.085	.065	3.60	3.86	1	10
Mid-Year DRA2 Independent Text Level	0	175	9.36	5.673	.429	8.51	10.21	1	28
	1	863	11.37	6.352	.216	10.95	11.79	1	34
	Total	1038	11.03	6.285	.195	10.65	11.41	1	34
Fall PALS Summed Score	0	172	52.61	16.347	1.246	50.15	55.07	3	88
	1	865	58.69	15.050	.512	57.68	59.69	9	90
	Total	1037	57.68	15.431	.479	56.74	58.62	3	90
Mid-Year PALS Summed Score	0	174	63.24	17.212	1.305	60.67	65.82	3	98
	1	870	68.59	15.881	.538	67.53	69.65	0	98
	Total	1044	67.70	16.225	.502	66.71	68.68	0	98
Fall PALS Spelling	0	172	15.30	8.830	.673	13.97	16.63	0	43
	1	865	18.95	9.313	.317	18.33	19.57	0	44
	Total	1037	18.34	9.329	.290	17.77	18.91	0	44
Mid-Year PALS Spelling	0	174	24.83	10.677	.809	23.23	26.43	0	52
	1	872	27.65	10.795	.366	26.93	28.37	0	52
	Total	1046	27.18	10.821	.335	26.52	27.84	0	52

Fall PALS Text Level

An independent-samples t test revealed that the average Fall PALS Text Level was significantly higher among the New Day Pre-K group (M = 3.86, SD = 2.10) than among the no Pre-K group (M = 3.07, SD = 1.87), t (1035) = 4.59, p <.001, Hedges g = .382. We can be 95% confident that the true difference between these means is CI = [.45, 1.13].

Mid-Year DRA2 Independent Text Level

An independent-samples t test revealed that the average Mid-Year DRA2 Independent Text Level was significantly higher among the New Day Pre-K group (M = 11.37, SD = 6.35) than among the no Pre-K group (M = 9.36, SD = 5.67), t (1036) = 3.88, p <.001, Hedges g = .32. We can be 95% confident that the true difference between these means is CI = [.99, 3.02].

Fall PALS Sum Score

An independent-samples t test revealed that the average Fall PALS Sum Score was significantly higher among the New Day Pre-K group (M = 58.69, SD = 15.05) than among the no Pre-K group (M = 52.61, SD = 16.34), t (1035) = 4.76, p <.001, Hedges g = .40. We can be 95% confident that the true difference between these means is CI = [3.57, 8.57].

Mid-Year PALS Sum Score

An independent-samples t test revealed that the average Mid-Year DRA2 Independent Text Level was significantly higher among the New Day Pre-K group (M = 68.59, SD = 15.88) than among the no Pre-K group (M = 63.24, SD = 17.21), t (1042) =

3.99, p <.001, Hedges g = .33. We can be 95% confident that the true difference

between these means is CI = [2.72, 7.97].

Fall PALS Spelling Score

An independent-samples t test revealed that the average Mid-Year DRA2

Independent Text Level was significantly higher among the New Day Pre-K group (M =

18.95, SD = 9.31) than among the no Pre-K group (M = 15.30, SD = 8.83), t (1035) =

4.72, p <.001, Hedges g = .40. We can be 95% confident that the true difference

between these means is CI = [2.13, 5.15].

Mid-Year PALS Spelling Score

An independent-samples t test revealed that the average Mid-Year DRA2

Independent Text Level was significantly higher among the New Day Pre-K group (M =

27.65, SD = 10.79) than among the no Pre-K group (M = 24.83, SD = 10.67), t (1044) =

3.15, p <.002, Hedges g = .26. We can be 95% confident that the true difference

between these means is CI = [1.06, 4.57].

Table 4.4 presents each of the independent-samples t test results.

Table 4.4

Reading Outcomes: Independent-Samples *t* Tests

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Fall PALS Text Level	EV assumed	5.012	.025	4.59	1035	.000	.793	.172	.454	1.131
	EV not assumed			4.97	264.52	.000	.793	.159	.479	1.107
Mid-Year DRA2 Independent Text Level	EV assumed	4.192	.041	3.88	1036	.000	2.010	.518	.994	3.025
	EV not assumed			4.18	270.17	.000	2.010	.480	1.064	2.955
Fall PALS Summed Score	EV assumed	1.290	.256	4.76	1035	.000	6.076	1.275	3.574	8.578
	EV not assumed			4.51	232.19	.000	6.076	1.347	3.422	8.731
Mid-Year PALS Summed Score	EV assumed	1.958	.162	3.99	1042	.000	5.347	1.338	2.722	7.972
	EV not assumed			3.78	235.56	.000	5.347	1.412	2.566	8.128
Fall PALS Spelling	EV assumed	1.012	.315	4.72	1035	.000	3.644	.771	2.132	5.157
	EV not assumed			4.89	252.56	.000	3.644	.744	2.179	5.110
Mid-Year PALS Spelling	EV assumed	.105	.746	3.15	1044	.002	2.820	.895	1.065	4.576
	EV not assumed			3.17	248.70	.002	2.820	.888	1.071	4.570

Effect Sizes

The effect size (ES) for each reading outcome was calculated using the mean difference divided by weighted and pooled standard deviation. Using the additional weight is recommended when the treatment and control groups are significantly different in size (Ellis, 2010), as is the case in this study. Cohen's *d* effect sizes are very similar, but lack the additional weight which is calculated by including the sample sizes of the control and treatment groups into the ES formula. The resulting formula produces the Hedges *g* ES (Formula 4.1), which is a *d*-based effect size (Ellis, 2010).

$$\textbf{Hedges' } g = \frac{M_1 - M_2}{SD*_{pooled}} \qquad \text{SD* pooled = weighted and pooled standard} \qquad (4.1)$$
$$\text{deviation.}$$

Ellis (2010)

The effect sizes were then translated into a percentile gain as a result of the treatment. For example, a .35 effect size translates into a 14% gain in achievement scores as a result of receiving the treatment. In other words, if a student scored in the 50th percentile (average) without the treatment, he or she would be expected to score in the 64th percentile after the treatment (above average). An effect size of .25 translates into a 10% gain as a result of treatment, for example, moving the student from the 50th to the 60th percentile. To put the following effect sizes into context, an effect size of .20 is a "reasonable minimal effect size level" and is still considered to be of practical significance (Lipsey, 1998). Table 4.5 illustrates the reading outcome effect sizes and the corresponding percentile gain and change for treatment students.

Table 4.5

Effect Sizes and Percentiles Gains for Reading Outcomes

	Reading Outcome	Effect Size	Percentile Gain	Percentile Change
1	Fall PALS Text Level	0.382	15	50th to 65th
2	Mid-Year DRA2 Text Level	0.321	13	50th to 63rd
3	Fall PALS Sum Score	0.398	16	50th to 66th
4	Mid-Year Sum Score	0.332	13	50th to 63rd
5	Fall PALS Spelling	0.395	16	50th to 66th
6	Mid-Year PALS Spelling	0.261	10	50th to 60th

These effect sizes are all slightly smaller than the Cohen's d effect sizes, because they take into account the difference in treatment and control group size and are based on a formula by Hedges (1981) that also removes a small positive bias in the Cohen's d. For example, the ES for the Falls PALS Text Level using Cohen's d is .397, while the same ES using the Hedges g formula is .382, as reported, representing a .015 smaller ES. Cohen's d would be appropriate to generalize results to the New Day City population, whereas to generalize results further, beyond the city, Hedges g is more appropriate.

Social / Behavioral Outcomes

Does attendance in New Day Pre-K make a sustained difference in children's social and emotional adjustment into first grade? Table 4.6 shows descriptive behavior statistics.

Behavior Sum Score

An independent-samples *t* test revealed that the average Behavior Sum Score was similar among the New Day Pre-K group (M = 2.04, SD = 3.66) and the no Pre-K group (M 2.13, SD = 3.96). See Table 4.7. In this case, the lower the score, the better the behavior.

Table 4.6

Behavior Sum Score: Descriptive Statistics by Group

	Treat	N	Mean	Std. Deviation	Std. Error Mean
Behavior Sum Score	1	866	2.04	3.660	.124
	0	173	2.13	3.960	.301

Table 4.7

Behavior Sum Score: Independent-Samples *t* Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. 2-tailed	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Behavior Sum Score	EV assumed	.58	.44	-.28	1037	.773	-.089	.309	-.696	.517
	EV not assumed			-.27	234.37	.785	-.089	.326	-.731	.553

Disaggregated Analysis for Subgroup Comparisons of Behavior Outcomes

School Quality & Behavior

The only subgroup where New Day Pre-K exerts a significant impact on the composite behavior outcome is fair quality schools. The SPSS file was split and separated by school quality in Kindergarten and an independent-samples t test was run. For the fair school quality subgroup, the treatment children have an average behavior composite score of 1.03 points less than the control (where 0 is perfect and a smaller score indicates better behavior) and the effect size is 0.49, a medium effect. A 0.49 effect size implies that New Day Pre-K produces a 19% gain for students in fair schools on overall behavior, as measured by a composite of seven citizenship skills. This would move a student from the 50th to the 69th percentile on the behavior sum score.

An independent-samples t test split by school quality revealed that only the fair quality schools average Behavior Sum Score was significantly better among the New Day Pre-K group (M = .87, SD = 1.75) than among the no Pre-K group (M = 1.90, SD = 3.09), t (94) = -1.96, p =.052, Hedges g = .49. We can be 95% confident that the true difference between these means is CI = [-2.07, .01]. (See Tables 4.8 and 4.9).

Table 4.8

Behavior Sum Score by School Quality: Group Statistics

School Quality in Kindergarten		Treat	N	Mean	Std. Deviation	Std. Error Mean
Excellent	Behavior Sum Score	1	42	1.17	2.219	.342
		0	3	.00	.000	.000
Good	Behavior Sum Score	1	124	1.96	3.771	.339
		0	26	2.46	3.787	.743
Fair	Behavior Sum Score	1	76	.87	1.746	.200
		0	20	1.90	3.093	.692
Low	Behavior Sum Score	1	321	2.46	4.060	.227
		0	71	2.38	4.291	.509
Very Low	Behavior Sum Score	1	298	2.04	3.635	.211
		0	53	1.83	4.023	.553

Table 4.9

Behavior Sum Score by School Quality: Independent-Samples *t* Test

School Quality in Kindergarten		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig	t	df	Sig. 2-tailed	Mean Diff.	Std. Error Diff.	95% Confidence Interval of the Difference Lower	Upper
Excellent Beh.	EV assumed	4.11	.05	.90	43	.373	1.16	1.29	-1.44	3.77
	EV not assumed			3.40	41.00	.001	1.16	.34	.47	1.85
Good Beh.	EV assumed	.045	.83	-.61	148	.538	-.50	.81	-2.11	1.10
	EV not assumed			-.61	36.15	.542	-.50	.81	-2.15	1.15
Fair Beh.	EV assumed	7.23	.01	-1.96	94	.052	-1.03	.52	-2.07	.01
	EV not assumed			-1.43	22.28	.166	-1.03	.72	-2.52	.46
Low Beh.	EV assumed	.33	.56	.15	390	.881	.08	.53	-.97	1.13
	EV not assumed			.14	99.62	.885	.08	.55	-1.02	1.18
Very Low Beh.	EV assumed	.09	.75	.38	349	.703	.21	.55	-.87	1.29
	EV not assumed			.35	67.94	.723	.21	.59	-.97	1.39

Summary

This study sought to answer two particular research questions: 1) does attending New Day Pre-K make a sustained impact on reading outcomes in first grade, and 2) does New Day Pre-K make a sustained impact on social and emotional outcomes in first grade? The propensity score matching method using optimal matching on a 1:5 fixed ratio was used to balance covariates between the control and treatment groups. After matching, the two groups had significantly balanced covariates on key variables such as

race, disability and income or SES, among others, as shown in the reduction of the standardized difference between the groups to the acceptable level of SD < 10% (actual average SD = 5.10).

An independent-samples t test was appropriate because two groups were being compared on a continuous outcome variable for each particular measure. Results indicated that attending New Day Pre-K had a significant positive effect on all six reading outcomes in first grade, including Fall and Mid-Year text level, spelling and sum score. The sum score is a composite of student's sight word reading, letter sound identification and spelling scores.

New Day Pre-K children appear to score similarly on report card citizenship skill results compared to the no Pre-K group. In a follow-up subgroup analysis of behavior outcomes by different school quality groups, there appears to be a correlation between attending a fair quality school in Kindergarten and having significantly better behavior for students who attended New Day Pre-K compared to those who did not receive any type of formal or institutional preschool.

CHAPTER 5

SUMMARY AND INTERPRETATION

Summary

Public Pre-K has grown substantially over the past fifteen years in the United States (Barnett, Carolan, Fitzgerald, & Squires, 2011) and is expected to continue expanding as the economy stabilizes. Comprehensive public preschool interventions targeting vulnerable children (Abecedarian, High/Scope Perry and Child-Parent Centers) have been found to make sustained long-term impacts for children (Reynolds, 2010) but scant research has investigated the longer-term impacts of attending typical public Pre-K programs, as opposed to Head Start, private preschool or intensive preschool models. Understanding the ability of typical Pre-K instruction to have sustained impacts into grade school is important because public Pre-K is now the largest provider of preschool slots in the county, and because Pre-K is still seen as a potential contributor to closing the achievement gap between at-risk and other children (Frede & Barnett, 2011). However, there is relatively little published research on K – 3 impacts as a result of Pre-K attendance. Studies that have investigated K – 3 impacts related to Pre-K attendance include Huang, Invernizzi & Drake (2011), Peisner-Feinburg, Elander & Maris, (2008), Magnuson, Ruhm & Waldfogel (2007a) and Frede (2009) among others. Further insight into the ability of Pre-K attendance to make a sustained difference for children into the early grade school years is needed to support efforts to create an aligned system of quality PK – 3 (Reynolds, Magnuson, & Ou, 2010).

Many challenges face researchers who are investigating the later impacts of Pre-K attendance. Reducing selection bias between the control and treatment groups remains an ongoing challenge, seen in studies where groups are imbalanced with respect to various risk factors (Xiang & Schweinhart, 2002). Reducing reliance on modeling assumptions by using sophisticated matching methods, such as propensity score matching, has been done only rarely (Magnuson, Ruhm & Waldfogel (2007a). Furthermore, very few longitudinal Pre-K studies in the United States have measured children's later behavior or social outcomes as a function of Pre-K attendance (Peisner-Feinburg, Elander & Maris, 2008).

Researchers investigating the later impacts of Pre-K attendance face other challenges as well. States have differing data collection systems (Hernandez, 2012) and different preschool standards. Even across districts in the same state there is no uniform Pre-K curriculum. Pre-K standards, data collection systems, funding arrangements, measures and curricula are not uniform across the country. However, realistic expectations for later Pre-K impacts can begin to emerge by conducting research on a wide variety Pre-K programs.

This study analyzed the effects of attending a school division Pre-K program in the southeastern United States. The division operates four independent early childhood centers serving approximately 2000 preschool children per year. The program is not in a state with universal public preschool access, but has created nearly universal access thanks to the creation of separate early childhood centers. The setting is primarily urban with a higher than average poverty rate (67% of all first graders live in low-income families). The division consists of a 30,000 students overall, drawn from a large and

diverse community. The city has a population of less than 250,000 people with two large racial groups: White and African American residents each account over 40% of the population. The Hispanic community is between five and 10% of the population.

New Day Pre-K is a grant-funded preschool program that is free for all children in the city. In 2012 – 2013, the Pre-K program was able to admit every child in the city who met age criteria, with the exception of just 30 students. In this sense, it is a nearly universal Pre-K program and places no restrictions on enrollment based on income or academic need, although children with greater academic need are admitted first. There is little external information available about the quality of the Pre-K experience in the city, although it meets all of the expected structural indicators of quality regarding teacher-child ratios, teacher early childhood credentials and use of state preschool objectives, among others. The particular curriculum being used when the first grade cohort went through the Pre-K program in 2010 – 2011 was based on the Harcourt Preschool Curriculum with a primary focus on early literacy. The adapted version of the curriculum used by the division was primarily teacher-centered with daily lesson plans provided to teachers revolving around literacy instruction. Math was taught for 2 – 3 days a week, while literacy received a daily focus.

The purpose of this study was to contribute to an empirical understanding of the relationship between district provided Pre-K and children's later reading and social outcomes by the beginning and middle of first grade. The study followed two main lines of inquiry: a) does attendance in New Day Pre-K make a sustained difference on reading achievement in first grade and b) does New Day Pre-K attendance make a sustained difference on children's later social and emotional adjustment into first grade?

To answer these questions, archived outcome data on the entire first grade cohort for the 2012 – 2013 year were taken from the PALS and DRA2 measures for reading achievement (text level, letter sound identification, sight words, spelling) and from report cards for citizenship (following the rules, responding appropriately to authority, working cooperatively with others, etc). This data represented the entire population of public first grade students in the city (n = 2221) and from this population a study sample was drawn which consisted of 1056 first graders, representing 176 control students who had never attended any type of formal or institutional preschool and 880 first graders who attended one of the early childhood centers in the city.

Propensity score matching was used to balance the covariates between the control and treatment groups. The particular matching method used was optimal matching on a 1:5 fixed ratio (176*5 = 880). Optimal matching on key covariates such as age, race, gender, disability and SES, successfully reduced the standardized difference of the covariates between the two groups to five percent, below the 10% threshold necessary for analysis (Rosenbaum & Rubin, 1983). The effects of various outcome measures were then analyzed using a series of independent-samples t tests. Significant findings were converted to effect sizes using a conservative version of the Cohen's d approach known as the Hedges g. Hedges g removes a slight positive bias present in the Cohen's d (Hedges, 1981) and is recommended for studies with unequal group size (Ellis, 2010). It is calculated using weighted and pooled standard deviation. Effect sizes were reported alongside the associated percentile gains expected as a result of attending the treatment. Where there was no statistically significant finding initially ($\alpha < .05$), further subgroup analysis was performed by school quality in kindergarten.

The study found significant effects of attending the New Day Pre-K, in the small to medium range, on all six reading measures in first grade. These included text level, spelling and a literacy sum score, taken in both the beginning and the middle of first grade. The average effect size across all literacy measures was .35, a small to moderate effect. There was no significant difference between New Day Pre-K attendees and the no preschool group on any measure of behavior, nor on the composite behavior sum score created for this study. Behavior results for the two groups were nearly identical. Further subgroup analysis by school quality showed that children who attended New Day Pre-K and went on to a "fair quality" school in kindergarten had significantly better behavior than children who had no preschool experience and who also attended the fair quality schools ($p = .052$, effect size .49).

Organization

This rest of the chapter is outlined as follows. First, a description and interpretation of each of the six reading outcomes is provided. Then, behavioral results are reviewed, followed by the corresponding subgroup analysis by school quality. After an interpretation of the findings, discussion on the meaning of the results for both reading and behavior is presented. Comments about the particular curriculum, instructional approaches and program are offered. Finally, implications for practice and policy are suggested and recommendations for future research are presented.

Interpretation

<u>Reading Achievement</u>

The influence of attending New Day Pre-K continues to have a significant effect on six measures of reading achievement taken in the beginning and middle of first grade. With effect sizes ranging from .26 - .40 for reading outcomes, first grade children in 2012 – 2013 who previously attended New Day Pre-K scored on average 10 – 19 % higher than the no preschool group.

The Fall PALS tests are administered by the teacher in late September and early October. The Fall PALS Text Level test is important because it determines for the teacher which child is at-risk of not meeting reading benchmarks. The original data was dummy coded as a continuous measure on a scale of 1 – 10, where 1 is called "readiness" and is the lowest reading level on the scale and 10 represents a fifth grade reading level. For example, if two children are tested using the Fall PALS Text Level measure and child A scores a 3 while child B scores a 4, the meaning is that child B is reading at one full text level above child A. Children who attended New Day Pre-K began first grade already reading nearly one full text level higher than the no preschool group. An independent-samples t test revealed that the average Fall PALS Text Level was significantly higher among the New Day Pre-K group (M = 3.86, SD = 2.10) than among the no Pre-K group (M = 3.07, SD = 1.87), t (1035) = 4.59, p <.001, Hedges g = .382. According to this measure, children from the New Day Pre-K group are less at risk for reading difficulty than the children who did not attend preschool.

To bring further meaning to these results, it is helpful to understand how the district uses the scores. Children entering first grade in New Day City are expected to score at a level 4 (which is named "pre-primer C" or simply "PPC" on the original assessment). In this case, we can see that the average reading score for all children who attended New Day Pre-K is 3.84, which means that the treatment group scores right around the expected benchmark for reading achievement in the beginning of first grade. However, the control group scores significantly lower: with an average score of 3.07. Children who did not attend preschool in New Day City are already reading nearly one level below the expected benchmark in the beginning of first grade.

Another way to consider the Fall PALS Text Level finding is to play the role of a first grade teacher working for New Day Schools. He or she will have children who attended the Pre-K program, nearly all of whom will begin first grade meeting the reading benchmark. This teacher will also have children who did not attend any type of preschool, most of whom will be reading below grade level to start the first grade year. After speaking with first grade teachers in this district, they confirmed that they could indeed identify who had attended New Day Pre-K and who did not, in general, simply by observing the reading differences among their students. This study brings empirical evidence to statements made by various first grade teachers that they could accurately "guess" who went to preschool and who did not. While the analysis shows that the difference between the groups on Fall PALS Text Level is statistically significant, and that the effect size is of medium strength at .38, it is perhaps more meaningful to consider that these scores translate into a reality in which one group of children is essentially on

grade level in reading versus another group who are nearly a full text level below the expected benchmark.

Given the primacy of the text reading level measure to reading performance scores, it is important to consider if and how the New Day Pre-K group differs from the no preschool group by the middle of first grade on a measure of their independent reading level. The Mid-Year DRA2 Text Level score is a continuous outcome measure on a dummy coded scale of 1 – 34. The real DRA2 test includes reading levels beyond level 34, but no first grade student in the city received a score higher than a 34 on the DRA2 test (n = 3) in the middle of first grade. As a reference point, a DRA2 34 is expected by the middle of third grade. This is not to say that there were no children who could read beyond a third grade level at the time, but rather that teachers simply did not test high-achieving first grade children to their capacity, responding to guidance from their administrators that it is sufficient to test children up to a particular point above grade level and then to discontinue testing in the interest of time, which is a common practice in the middle of the year. Nevertheless, it is unlikely that this truncated testing for high-achievers would significantly alter the average scores between the treatment and control groups in this study.

To help put the Mid-Year DRA2 Text Level scores in context, a DRA2 score of 10 – 12 was expected for the middle of first grade by the division. The PALS Text Level test is not administered in the middle of the year. As an independent reading test, the DRA2 is slightly different from the PALS Text Level test in that the PALS test is used as a screening tool, to screen children who are at risk of reading difficulty, whereas the DRA2 test is designed to identify a child's true independent reading level. For reference,

the expected rate of growth and corresponding DRA2 scores for a child in the first grade

is illustrated in the DRA2 Reading Progress Chart, Table 5.1.

Table 5.1

DRA2 Reading Progress Chart for First Grade

September	Level 4 - 6
January	Level 10 - 12
May - June	Level 16 - 18

The reading progress chart used by the division helps interpret the meaning of the

results reported in chapter 4 regarding the Mid-Year DRA2 Independent Text Level

score. Using this chart we would expect to see most children scoring in the DRA2 10 –

12 range at the mid-year point. An independent-samples t test revealed that the average

Mid-Year DRA2 Independent Text Level was significantly higher among the New Day

Pre-K group (M = 11.37, SD = 6.35) than among the no Pre-K group (M = 9.36, SD =

5.67), t (1036) = 3.88, p <.001, Hedges g = .32. But perhaps more important, again, is to

consider reality for the no preschool group. Just like the beginning of first grade, the no

preschool group is still reading below the expected benchmark (DRA2 10), while the

New Day Pre-K group is scoring right in the expected range at 11.37, and is even on the

positive side of the range with an average score above 11.00.

There is continuity from the beginning to the middle of first grade, in both groups

of students: the New Day Pre-K group is remaining on grade level on measures of text

level reading and the no preschool group is remaining below grade level. This does not imply that the no preschool group is not making progress over time, but merely states that they entered first grade below grade level and continue to be below grade level by the middle of the year.

It is necessary to interpret these reading text level findings within reason: using benchmarks to help interpret student test results often overlooks the amount of growth a child has achieved. Promoting benchmark evaluation is not the intention of this narrative. To say that a child is below benchmark is meaningless if we do not consider that child's starting and ending scores in relation to other children: he or she may have made great progress but started much lower than another child, and would still remain "below" benchmark while a higher achieving child may have made less progress over the same time but still ends on or above the benchmark. However, the findings of this study reveals a pattern: the New Day Pre-K group is generally on "grade level" from the beginning to the middle of first grade while the no preschool group is a full text level behind the treatment group at both time points, and achieves below the division benchmarks for text level reading as well.

The next reading measures to analyze are the Fall and Mid-Year Sum Scores. The Sum Score is a district-created composite that is the sum of the child's spelling, sight word recognition, and letter sound identification scores. The Sum Score is totaled by adding all the points awarded from the spelling, word list and letter sound tests. The Sum Score is a useful measure to consider the total non-text reading abilities of a child. That is to say, the Sum Score totals all reading results into one score, excluding the text level result that was described above. If a child scores higher than another on the Sum Score it

can be assumed that he or she has better early literacy skills overall. Some children may score a bit lower or higher than others on the sight word test or the spelling test, so the Sum Score is used to help even out these differences and identify more clearly what a child can do overall, excluding text reading ability. The New Day Pre-K group scored significantly higher than the no preschool group on the Fall PALS Sum Score (M = 58.69 versus M = 52.61). This is a 6.05 point average advantage and carries an effect size of .40, a medium effect. A perfect score on the Sum Score is 90 points (44 points from spelling, plus 26 letter sounds, plus 20 sight words correctly read). Sixty-five percent of all children scored in the 50 – 65 point range on the Sum Score. These results indicate that New Day Pre-K children have significantly greater early phonological and sight word recognition skills than the no preschool children, and that the magnitude of the effect is large enough to be of practical relevance, meaning that the differences are apparent in reality when interacting with the two groups of children.

The Mid-Year PALS Sum Score results are also significantly different between the New Day Pre-K and no preschool groups (M = 68.59 vs. M = 63.24, respectively). The strength of the effect is slightly smaller at .33, though still educationally relevant. These results mean that children who attended the division Pre-K program have better combined phonological and sight word recognition skills than children who did not attend any type of preschool. However, it can also be seen that the strength of the effect of attending Pre-K on Sum Score results was slightly greater in the beginning of first grade compared to the middle of the year. The slightly reduced effect size on the Sum Score from the beginning to the middle of the year is too small to interpret with confidence. However, there is a clear trend across all reading measures showing that the strength of

the effect is consistently greater in the beginning of the year, and consistently drops slightly by the middle of the year, revealing a systematic pattern that the effect of Pre-K attendance appears to weaken slightly, though still significant, as students progress from the beginning to the middle of the first grade year. However, this change over time was not a direct research question in this study because the three months separating the two testing windows was too short to confidently interpret a difference. It would be more meaningful, for example, to compare reading results from the beginning of the year to the end of the year, or from first grade to second grade, but this data was not available at the time of collection.

The final set of literacy measures to interpret are the spelling results from the beginning and the middle of the first grade as a function of attending or not attending New Day Pre-K. The difference between the treatment and control groups on the Fall PALS Spelling test was significant (18.95 vs. 15.30 , respectively, for a difference of plus 3.65 points on average) with a medium effect size of .40. (Again, all of the effect sizes reported in this study are calculated using the Hedges g formula which is slightly more conservative than that the Cohen's d.)

To understand the Fall PALS Spelling Test results it is helpful to consider the nature of the spelling test that children had to take, and the way it was scored. Children were verbally presented with twenty words to spell. The word was stated by the teacher, used in a sentence and re-stated, and then children wrote the word on a scoring sheet. Scoring was not based on one point for each word. Rather, each feature within the word was awarded 1 point, and the word itself was awarded 1 point. So the word "chin" would be scored as 1 point for the "ch" and 1 point for the entire word. Spelling "chin" correctly

gave the child 2 points. Three letter words such as "fit" would be scored as 1 point for the middle vowel "i" or 1 point for the beginning and ending sounds (f / t) and a point for the entire word. Based on this scoring method, the highest score a child could receive on the spelling test, for spelling all 20 words correctly, was 44 points for the beginning of the year, and 52 points for spelling 20 words correctly on the mid-year test. The higher possible points awarded on the mid-year spelling test is indicative of giving children slightly more complicated words, with additional within-word features to spell correctly.

Table 5.2 shows the mean and median score for all children combined and the cut points for students scoring at the 20, 40, 60 or 80th percentiles. The bottom 20% of students scored at or below 10 points on the Fall PALS Spelling Test, which means that they scored about five words correctly on the twenty word test. Half scored below 17 on the same test, which is equivalent to spelling about 8 words correctly out of the twenty. At the 60th percentile, students were spelling about half of the words correctly and at the 80th percentile, students were spelling about 13 of 20 words correctly. Considering that the median score was 17 points, we can see that the New Day Pre-K group's average Fall PALS Spelling Test score fell in the top 50th percentile (M = 18.95), while the no preschool groups' spelling scores fell below the 50th percentile (M = 15.30).

Table 5.2

Fall PALS Spelling Statistics

	Valid	1037
Mean		18.34
Median		17.00
Std. Deviation		9.329
Percentiles	20th	10.00
	40th	15.00
	60th	20.00
	80th	26.00

The statistical output, above, reports a mean spelling score for all students in the sample of 18.34, which is misleading because this study used a 1:5 treatment to control ratio, meaning that the mean score is skewed in favor of the treatment cases: five times more treatment cases are being added into the calculation of this mean score. The "true" average between the two groups is to add the average of the two groups together and divide by two, which yields a corrected study mean of 17.125. When considering either the median score of 17 or the corrected study mean of 17.125, it is clear that the no preschool group is performing below average on the spelling assessment. Another factor in interpreting these spelling results is to consider how high the New Day Pre-K group scored relative to the percentile cut points presented in the statistics output above. A descriptive frequency run in SPSS reveals that a score of 18.95, rounded to 19, places the New Day Pre-K group in the 59th percentile. The no preschool groups' Fall spelling results fall in the 41st percentile. The true difference between the two groups in percentile gain terms as a function of Pre-K attendance is 18 percentage points, which is slightly

higher than the 16th percentile gain predicted by the .40 effect size. Regardless of how the results are interpreted, New Day Pre-K continues to make a meaningful contribution to students' first grade spelling achievement at the beginning of year after balancing the groups on key covariates such as age, race, gender, disability and socio-economic status.

The second measure of students' spelling achievement was taken in the middle of first grade. Like the beginning of the year, New Day Pre-K students continued to score significantly higher than the no preschool group, but with a tighter range (2.82 points apart at mid-year compared to 3.65 points apart in the beginning of the year) resulting in a reduced effect size of .26, which in reality translated to a percentile gain of 11% for the treatment group over the no preschool group. Interestingly, the no preschool group improved their results slightly from the 41st to the 45th percentile from the beginning to the middle of the year in spelling, while the New Day Pre-K group's percentile results dropped over the same time period, from the 59th to the 56th percentile. With a sample size of just two tests, it is hard to read into these results to say with confidence if the no preschool group is gradually catching up to the treatment group or if the treatment group is simply not progressing as rapidly as before. The spelling results appear to suggest that both trends are happening simultaneously: the no preschool group is starting to improve overall and the New Day Pre-K group is not sustaining the same degree of advantage from the beginning to the middle of the year.

Discussing the degree of change from the beginning to the middle of the year may be speculative because we do not know if the harder spelling test taken in the middle of the year impacted the ability of students to sustain higher scores. Also, a single test of twenty words is a small sample size to judge student abilities. The division emphasized

small group instruction using word study activities to improve spelling during the 2010 academic year in first grade, with an increased focus on struggling students, providing them with additional intervention. It is possible, therefore, that these additional efforts assisted the control group more than the New Day Pre-K group during the first four months of the first grade year. Without considering the degree of change from the beginning to the middle of first grade, the independent-samples t test still shows a significant difference at the p < .002 level between the New Day Pre-K and control groups on the Mid-Year PALS Spelling Test.

Teacher Reported Citizenship Behavior

This study used teacher-report data in the form of report card results on seven citizenship skills to create a behavior composite index score. There was no systematic pattern behind how teachers evaluated children on any particular citizenship skill, except for "follows the rules." Teachers were twice as strict to grade a child down on following the rules, than on any other measure of citizenship. The rest of the six citizenship skill grades seem to have been graded fairly randomly, but with even results across the spectrum, meaning that any particular grouping or combination of the skills would make no difference in analysis. Due to this subjective and unsystematic grading pattern across the seven skill areas, it was decided the best approach would be to total all seven skill areas and award each child an aggregated behavior score. It could be argued that mixing and matching particular groups of citizenship skills for analysis would be revealing, but the random assignment of grades across the spectrum of skills makes that possibility

unlikely. Further complicating this situation is the fact that the children did not have the same report card for Pre-K as they did in first grade, meaning that there is not a true baseline behavioral score to consider when interpreting the findings. The situation is slightly different for literacy, because it can be assumed that just about all children did not know how to read when they were beginning preschool or were at that age, so a "baseline" can be estimated. However, children come into preschool already behaving and interacting at a certain level, making any findings even harder to relate to the Pre-K experience or lack therefore. Nevertheless, the groups were significantly balanced after matching to permit analysis on later behavioral outcomes as a function of attending Pre-K, which was important for several reasons: a) assessing possible behavior effects recognizes that cognitive and affective development are interrelated and equally important to the promotional of holistic child development, b) results can assist the school division to examine their approaches to social and emotional development, and c) the influence of typical Pre-K on children's K - 3 behavior is not well understood or studied in the Pre-K evaluation literature.

An independent-samples *t* test was run on all seven citizenship skills, and no result was significant. This study, however, only reports the results of the behavior sum score for the reasons mentioned above: there was no systematic pattern to how teachers chose to grade a child down on one skill versus another, and therefore looking at the individual citizenship skill results was inappropriate. The least biased approach to analyzing the behavior grades was to create a scaled index for the total behavior score. To review the results, the New Day Pre-K group had an average Behavior Sum Score of 2.04 versus the no preschool group of 2.13, where 0 is a perfect score (all Satisfactory, S,

marks for all seven citizenship skills). Two marking periods of grades were summarized to create the Behavior Sum Score. Marking period 1 and marking period 2 were added together. Thus, a total of 14 unique grades were assigned to each child in this study. Therefore, a score of 2.00 means that the child earned 6 S's and 1 N (needs improvement) in marking period one, and 6 S's and 1 N in marking period two. Or, in other words, a score of 2.00 means that the child earned all S's and 1 N on each report card for citizenship skills. This study is reporting the summed results of two marking periods together, which is why the average behavior sum score is around 2.00 rather than 1.00. Analysis revealed that the New Day Pre-K group at least did not perform worse on the behavior sum score than the no preschool group. This is noteworthy to mention because some prior research shows that low quality preschool has a detrimental effect on children's later behavior (Sylva, 2011).

Interestingly, before improving the covariate balance between the groups using optimal matching on the propensity score, preliminary analysis had shown that attending New Day Pre-K made a significant difference on Following the Rules when compared to the no preschool group. The treatment group scored significantly better on Rule Following before covariate balancing. However, after optimal matching to balance the covariates across the groups no difference was found in follow-up tests of significance on Following the Rules, or on any other particular citizenship skill, or on the behavior sum score.

<u>Taking School Quality into Account</u>

When data were disaggregated by school quality, it was found that the New Day Pre-K students had significantly better behavior than the no preschool children among children who attended fair quality schools ($p = .052$). No other school quality subgroup made a significant difference on the behavioral outcomes of children in first grade as a function of attending New Day Pre-K. Does the combination of fair quality schools and attending New Day Pre-K really result in improved behavior in first grade? Future research would need to determine what other factors may be involved in this finding and why this may be the case. In this particular study, the fair quality subgroup had 76 treatment cases and 20 control cases, which is a very small control group to use for making any meaningful conclusions about the combination of fair quality schools and New Day Pre-K attendance on behavior by first grade.

Discussion

<u>Reading</u>

This study found that students who attended the New Day Pre-K program in one of the four early childhood centers administered by the school division performed significantly higher across three different types of literacy measures in both the beginning and the middle of first grade. There was also an observable trend that the effect sizes across all literacy measures were higher at the beginning of the year than in the middle of

the year, for all test reading results. By considering all three of these literacy measures together (text level reading ability, spelling, and the sum score of spelling, sight words and letter sound identification) a more complete picture of a child's reading ability begins to emerge. When considering all of the literacy results together, it is clear that the New Day Pre-K program has a meaningful impact on children into first grade compared to those children who attended no preschool. Even the smallest effect size found in this study (0.26 for Mid-Year Spelling) is still consequential as an indicator that the treatment is effective when compared to no treatment (Lipsey, 1998).

Are the literacy results surprising? When considering the focused approach taken to teaching literacy in the New Day Pre-K program, described in more detail later, these results are perhaps less surprising. What is perhaps more encouraging is the strength of the effect on literacy overall as a result of attending the cities' Pre-K program. The average effect size across all six measures of reading and early literacy achievement is .35 which translates into a 14% gain for the Pre-K group over the no preschool group.

It is well established in the literature that attending Pre-K increases students scores on school readiness measures for literacy and math (Wong, Cook, Barnett & Jung, 2008). What has been less well studied is the degree to which these gains can be sustained into early grade school. This has been of interest given the experience of particular intensive public preschool programs (Abecedarian, High/Scope Perry, Child-Parent Centers) to foster long-term impacts in children that can last for many additional years, even into adulthood (Reynolds, 2010). What remains unknown is the degree to which typically operating public Pre-K can raise a child's trajectory of learning into later grade school years, and the particular factors that may be necessary to sustain those gains.

This study suggests that New Day Pre-K can significantly elevate a child's literacy achievement with meaningful effects sustained through the middle of first grade.

Furthermore, some studies on the longer-term impact of Pre-K have lacked a proper matching method creating doubt that the control and treatment groups are balanced. By using a more rigorous matching method to balance covariates, this study improved the quality of the evidence behind the sustainability of literacy impacts into first grade as a function of New Day Pre-K attendance. Also, by choosing a more conservative formula for calculating the effect size, these results are more likely to generalize to a wider sample beyond the population of New Day City. Before making such an assumption, however, the consumer of the research would first need to consider the demographic characteristics of New Day City in relation to any other. The extent to which the literacy gains of the New Day Pre-K group can be sustained beyond first grade remains unknown.

With regards to the quality of the New Day Pre-K program, all that can be said is that this program had sufficient structural quality based on such indicators as teacher certification in early childhood education, teacher-child ratios, presence of assistant teachers, the use of a formal literacy-based curriculum and adoption of state preschool objectives for learning. Some would argue that the improved literacy scores sustained into first grade is evidence for the effectiveness of the Pre-K program, however, no judgment is being passed in this study on the true quality of New Day Pre-K.

<u>Social and Emotional Outcomes</u>

Some policy and decision-makers hold the view that the primary motivation for creating a more universal model of Pre-K education is to develop key literacy and math skills in young children. This view tends not to value improved social and emotional adjustment as a significant justification for funding Pre-K. As a result, little systematic instruction is given to children with regards to their social and emotional development across the early years. The view is that as long as behavior is functionally reasonable, children can successfully move through the schooling process to allow the "important" work of academic instruction to take place. Meanwhile, particularly "problematic" children are put on special behavior plans, sent to school guidance counselors, moved to a "better" teacher's classroom, referred for other services, or punished on an ongoing basis in the form of suspensions and other consequences. This assortment of approaches to "helping" improve social and emotional growth and development appears to be management-based, to simply manage difficult behavior, rather than transformative. In general, there does not appear to be any purposeful or scientific approach being applied across early school years to significantly change most children's social and emotional abilities and understandings. Some may go as far as to say that changing social and emotional abilities in children is not the purpose of American education and / or that it cannot be achieved.

This discussion about the wider absence of explicit instruction for social and emotional development helps place the insignificant behavioral findings of this study into context. Without an intentional focus on social and emotional development that is

reflected in the daily activities of Pre-K, it is unscientific to assume that such a preschool experience would cause children to "behave" better than non preschool children in later years. In other words, with no further instructional input for social adjustment by the preschool program, it would be expected that both the preschool and non preschool groups would exhibit the same general citizenship results on report cards in first grade. The two groups had nearly identical behavior sum scores ($M = 2.13$ versus $M = 2.04$). Therefore, the result of this study that the New Day Pre-K and the no preschool group scored similarly on a composite behavior index is unsurprising and expected. Additional teacher training, or lesson plan support, was not provided to emphasize the teaching of social or behavioral skills to children.

In essence, teachers did not have a systematic approach to boosting student social and emotional development built into the Pre-K program (unlike literacy instruction). When this lack of a systematic approach to affective development was explored further, it was found that holding an early childhood teaching credential is assumed to provide teachers with some degree of expertise in developmentally appropriate practice, broadly, that may translate into some form of sensitivity for children's unique affective development. This idea that an early childhood credential "covers" the subject of social and emotional development sufficiently to translate into practices in the classroom is speculative. There is no empirical evidence in this study to suggest that a teachers' early childhood credential translates to meaningful or sustained social development activities in the classroom that would yield longer-term effects. In reality, some form of school-wide support for any particular practice is required to *broadly* implement any educational intervention. Future mixed-methods research would be needed to investigate these ideas.

Before concluding that New Day Pre-K made no difference on children's later behavior, it is important to acknowledge that the program did not try to significantly change children's social or behavioral interactions and had very little, if any, curricular activities designed explicitly for this purpose. Finally, it should be acknowledged that there is not necessarily a scientific connection between a child's social and emotional development and the seven citizenship skills being scored on the report card. The report card was a convenient measure for behavior but is not be the most accurate, either.

Other Subjects: Besides Early Literacy

Unfortunately, this study was not able to gather math outcome data to test the possible sustained effect of Pre-K attendance on first grade math achievement in the city. However, the New Day Pre-K program operating in 2010 had been designed to boost early literacy, not math, as its primary instructional objective.

An examination of the New Day Pre-K curriculum used in 2010 revealed a more teacher-centered and direct instruction model for literacy, with only sparse attention to math, science or other subjects. Sparse attention to other subjects means that math and science centers and stations were created for children, but intentional instruction in these or other subjects was much less consistent compared to the daily literacy focus. For example, in discussions it was found that math was taught just three days a week at best in 2010 while literacy was taught on a daily basis.

Curriculum Used by New Day Pre-K

The New Day Pre-K program used an adapted version of the Harcourt Preschool Curriculum, which is a literacy-focused curriculum. While the curriculum was not scripted in terms of the exact language a teacher should speak for every given activity, it included complete lesson plans, materials and activities outlined for each day of the week revolving around early literacy skills. The particular approach to teaching phonological awareness and early literacy was a more formal or traditional model as compared to other constructivist, and child-centered approaches. The attitude of more regimented curricular approaches tends to focus on skill development rather than creating learners who both love the subject and see themselves as writers, readers and thinkers, for example. Despite all these considerations, the objective of the New Day Pre-K curriculum and program at the time to emphasize early literacy did result in children who continued to outperform the no preschool group on early literacy and reading skills into the middle of first grade.

Implications and Recommendations

For Practitioners and School Divisions

Teachers will want to consider these findings from several perspectives. From the perspective of test scores, the results suggest that direct instruction in early literacy provided in preschool continues to make a sustained difference in children's first grade reading achievement. However, this does not automatically imply that the particular

instructional methods used to achieve this result are in-line with child-centered practice or that future results could not be improved upon even further. The results of this study can show to teachers that there is a value in improving literacy instruction at the preschool level, but this focus should not come at the expense of other cognitive developments such as mathematical thinking and exploratory engagements with scientific questions.

Teachers and instructional leaders also need to consider the finding in this study that children's behavior in first grade was unaffected by participation in the Pre-K program. High quality preschool has been associated with improved behavior outcomes in later years in other settings (Sylva, 2011). If a Pre-K program does not also improve a child's citizenship behavior, can it be said to be truly of "high quality?" Many early childhood educators believe that children need an experience that is balanced affectively and cognitively. The challenge to educators in similar school divisions is to experiment with how to improve young children's social and emotional development along with their cognitive development, and not to settle for a literacy-only view of quality when implementing or reforming Pre-K.

Policy Recommendations

The policy implications of this study are necessarily limited due to the small sample size and the placement of the study in a single city. Nevertheless, many of the characteristics of the New Day Pre-K program are common, particularly the emphasis on

early literacy, the larger scale mixed urban population, and the high concentration of poverty in the region.

Many decision makers are questioning the efficacy of providing universal public Pre-K through the public school system. Challenges with finding space and resources are major obstacles. However, the experience in New Day shows that a single city can expand their slots to provide nearly universal access even when the state as a whole does not provide universal preschool. It is far beyond the scope of this study to propose that universal preschool should or should not be offered, but this study does illustrate an example of how to solve some of the problems of space and limited resources such that nearly all children are accepted into the program while also achieving literacy results that are sustained into first grade. It is recommended that future approaches to early education re-emphasize a balanced view of development, because there does not seem to be a benefit to children's social and emotional well-being from attending this version of typical Pre-K.

Recommendations for Future Research

Longitudinal research is just beginning to address questions of the efficacy of typically functioning public Pre-K in the United States. So far, the literature has been mostly focused on later literacy and math outcomes, with just a handful of studies including measures for later behavior. Future research using non-biased behavioral measures is needed to determine if Pre-K can make a positive longer-term impact on children's social development. This particular study can also be replicated in coming

years by the division in question as a way of determining if and how their reforms at the Pre-K level have been effective. Future research needs to include more time points along the K – 3 continuum to bring more insight into the question of fading effects. With a short time point between the beginning and the middle of first grade, this study was not able to provide enough information about various changes over time, and future studies should compare results from the beginning of first grade to the end of first grade and into second and third grade.

Conclusion

It can be said, at a minimum, that despite the philosophical differences held by various educators as to how to create child-friendly preschool instruction, that is research-based and developmentally appropriate and attentive to the holistic needs of the child, nevertheless, the instruction provided in the New Day Pre-K program in 2010 did result in sustained, improved reading achievement into first grade without worsening children's traditionally viewed citizenship behaviors.

REFERENCES

Administration of Children and Families. (2013). *History of Head Start.* U.S. Department

of Health and Human Services, Office of Head Start.

Retrieved from http://www.acf.hhs.gov/programs/ohs/about/history-of-head-start

Aguilar, R. & Tansini, R. (2011). Joint analysis of preschool attendance and school

performance in the short and long-run. *International Journal of Educational*

Development, 32(2), 224 - 231.

Barnett, W. S., Carolan, M. E., Fitzgerald, J. & Squires, J. H. (2011). The state of

preschool 2011: State preschool yearbook. *New Brunswick, NJ: National Institute for*

Early Education Research.

Barnett, W. S. & Frede, E. (2010). The Promise of Preschool. *American Educator,* 34(1),

21-40.

Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R. & Burns,

S. (2008). Educational effects of the Tools of the Mind curriculum: A randomized

trial. *Early Childhood Research Quarterly,* 23(3), 299–313.

Berlinski, S., Galiani, S. & Gertler, P. (2009). The effect of pre-primary education on

primary school performance. *Journal of Public Economics,* 93(1-2), 219–234.

Bredekamp, S., & Copple, C. (Eds.). (1997). Developmentally appropriate practice in

early childhood programs (Rev. ed.). Washington, D.C: NAEYC.

Brooks-Gunn, J., Duncan, G. J., Klebanov, P. K. & Sealand, N. (1993). Do

neighborhoods influence child and adolescent development? *American journal of*

sociology, 353–395.

Camilli, G., Vargas, S., Ryan, S. & Barnett, W. S. (2010). Meta-analysis of the effects of

early education interventions on cognitive and social development. *The Teachers*

College Record, 112(3).

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.

Cauthen, N. K., & Fass, S. (2008). Measuring income and poverty in the United States.

New York, NY: National Center for Children in Poverty, Columbia University,

Mailman School of Public Health

Clements, D. H. & Sarama, J. (2007). Effects of a preschool mathematics curriculum:

Summative research on the Building Blocks project. *Journal for Research in*

Mathematics Education, 136–163.

Clifford, R., Reszka, S. (2010). Reliability and validity of the early childhood

environment rating scale. Working paper. *FPG Child Development Institute.*

Retrieved from http://www.ersi.info/PDF/ReliabilityEcers.pdf

Copple, C., & Bredekamp, S., (Eds.). (2008). Developmentally appropriate practice in

early childhood programs (Third ed.). Washington, D.C: NAEYC.

Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal,

M., Bryant, D., et al. (2009). The relations of observed pre-K classroom quality

profiles to children's achievement and social competence. *Early Education and*

Development, 20(2), 346–372. doi:10.1080/10409280802581284

Currie, J. & Thomas, D. (2000). School quality and the longer-term effects of Head Start,

34(4), 755–774.

D'Onise, K., Lynch, J. W., Sawyer, M. G. & McDermott, R. A. (2010). Can preschool improve child health outcomes? A systematic review. *Social Science & Medicine*, *70*(9), 1423–1440.

Douglas-Hall, A., Chau, M., Koball, H. (2006) *Basic Facts About Low-Income Children Birth to Age 6*, National Center for Children in Poverty Retrieved from http://www.nccp.org/publications/pub_680.html

Ellis, P.D. (2010), "Effect Size FAQs," retrieved on June 26, 2013 from *http://www.effectsizefaq.com*

Falk, B. (2012). *Defending Childhood: Keeping the Promise of Early Education*. Teachers College Press.

Frede, E., & Barnett, W. S. (2011). Why Pre-K Is critical to closing the achievement gap. *Principal*, *90*(5), 8-11.

Frede, E. C., Jung, K., Barnett, W. S., Figueras, A., (US), N. I. for E. E. R. & of Education, N. J. D. (2009). *The APPLES Blossom: Abbott Preschool Program Longitudinal Effects Study (APPLES): Preliminary Results Through 2nd Grade: Interim Report*. National Institute for 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 for Early Education Research*.

Gilliam, W. S. & Zigler, E. F. (2001). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly*, *15*(4), 441–473.

Gormley Jr, W. T., Gayer, T., Phillips, D. & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41*(6), 872. doi:10.1037/0012-1649.41.6.872

Gu, X. S., and Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2(4), 405-420.

Heckman, J. J. & Masterov, D. V. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy, 29*(3), 446–493. doi:10.1111/j.1467-9353.2007.00359.x

HEAD START ACT as amended. 42 USC 9801. (2007).

Hedges, L.V. (1981), "Distribution theory for Glass's estimator of effect size and related estimators," *Journal of Educational Statistics*, 6(2): 106-128.

Henry, G. T., Gordon, C. S., Henderson, L. W. & Ponder, B. D. (2003). Georgia pre-k longitudinal study: Final report 1996-2001. *Atlanta, GA: Georgia State University, Andrew Young School of Policy Studies.*

Hernandez, D. J. (2012). *PreK-3rd: Next Steps for State Longitudinal Data Systems. Policy to Action Brief (8).* Foundation for Child Development.

Hernandez, D. J. & Foundation, A. E. C. (2012). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation (updated).* The Annie E. Casey Foundation.

Ho, D.E., Imai, K., King, G., Stuart, E.A., (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42*(8), 1 - 28. Online: http://www.jstatsoft.org/v42/i08/

Huang, F. L., Invernizzi, M. A. & Drake, E. A. (2011). The differential effects of preschool: Evidence from Virginia. *Early Childhood Research Quarterly.* doi:10.1016/j.ecresq.2011.03.006

Hustedt, J. T., Barnett, W. S. & Jung, K. (2008). Longitudinal Effects of the Arkansas Better Chance Program: Findings from Kindergarten and First Grade.New Brunswick, NJ: National Institute for Early Education Research, Rutgers University

Hustedt, J. T., Barnett, W. S., Jung, K. & Figueras-Daniel, A. (2009). Continued Impacts of New Mexico PreK on Children's Readiness for Kindergarten.*New Brunswick, NJ: National Institute for Early Education Research.*

Knudsen, E. I., Heckman, J. J., Cameron, J. L. & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences, 103*(27), 10155–10162.

Lee, V. E. & Loeb, S. (1995). Where do Head Start attendees end up? One reason why preschool effects fade out. *Educational evaluation and policy analysis, 17*(1), 62–82.

Lipsey, M. W. (1998). Design sensitivity: Statistical power for applied experimental research. In L. Bickman & D. J. Rog (Eds.), Handbook of applied social research methods (pp. 39-68). Thousand Oaks, CA: Sage.

Lipsey, M. W., Farran, D. C., Bilbrey, C., Hofer, K. G. & Dong, N. (2011). Initial Results of the Evaluation of the Tennessee Voluntary Pre-K Program. *Peabody Research Institute, Vanderbilt University, Nashville, TN.*

Loveless, T., Parkas, S. & Duffett, A. (2008). High-Achieving Students in the Era of NCLB. *Thomas B. Fordham Institute*, 84.

Luthar, S. S. (2003). *Resilience and vulnerability: Adaptation in the context of childhood adversities*. Cambridge University Press.

Ma, X., Shen, J., Kavanaugh, A., Lu, X., Brandi, K., Goodman, J., Till, L., et al. (2011). Effects of Quality Improvement System for Child Care Centers. *Journal of Research in Childhood Education*, *25*(4), 399–414.

Magnuson, K. A., Ruhm, C. & Waldfogel, J. (2007a). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly*, *22*(1), 18–38.

Magnuson, K. A., Ruhm, C. & Waldfogel, J. (2007b). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*(1), 33–51. doi:10.1016/j.econedurev.2005.09.008

Maloffeva, E., Daniel-Echols, M. & Xiang, Z. (2007). Findings from the Michigan School Readiness program 6 to 8 follow up study. *High/Scope Educational Research Foundation*.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child development*, *79*(3), 732–749. doi:10.1111/j.1467-8624.2008.01154.x

Nelson, G., Westhues, A. & MacLeod, J. (2003). A Meta-Analysis of Longitudinal Research on Preschool Prevention Programs for Children. *Prevention & Treatment*, *6*(1), 31a.

Peisner-Feinberg, E. S., Elander, K. & Maris, C. L. (2008). Evaluation of the North Carolina More at Four Pre-kindergarten Program Year 6 Report: Children's

Longitudinal Outcomes and Program Quality over Time (2003-2007).Chapel Hill, NC: FPG Child Development Institute.

Phillipsen, L. C., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly, 12*, 281-303.

Pianta, R. C., Barnett, W. S., Burchinal, M. & Thornburg, K. R. (2009). The Effects of Preschool Education What We Know, How Public Policy Is or Is Not Aligned With the Evidence Base, and What We Need to Know. *Psychological Science in the Public Interest, 10*(2), 49–88. doi:10.1177/1529100610381908

Pianta, R., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Baltimore, MD: Brookes, Inc.

Reynolds, A. J., Magnuson, K. A. & Ou, S. R. (2010). Preschool-to-third grade programs and practices: A review of research. *Children and Youth Services Review, 32*(8), 1121–1131. doi:10.1016/j.childyouth.2009.10.017

Riddle, W. (2012). Major Accountability Themes of Second-Round State Applications for NCLB Waivers. *Center on Education Policy*.

Rosenbaum, P. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 604–620.

Schweinhart, L. J. & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool Study through age 40*. Ypsilanti, MI: High/Scope Press.

Stuart, E. & Rubin, D. B. (2004). Matching methods for causal inference: Designing observational studies. *Harvard University Department of Statistics*.

Stuart, E.A. (2010). Matching methods for causal inference: A review and a look

   forward. *Statistical Science, 25* (1), 1 - 21. doi: 10.1214/09-STS313

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. & Taggart, B. (2004). The

   effective provision of pre-school education (EPPE) project: Final Report: A

   longitudinal study funded by the DfES 1997-2004. University of London, Institute of

   Education.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. & Taggart, B. (2011). Pre-

   school quality and educational outcomes at age 11: Low quality has little benefit.

   Journal of Early Childhood Research, 9(2), 109–124.

   doi:10.1177/1476718X10387900

Temple, J., Arteaga, I. & Reynolds, A. (2010). Longer-term effects of preschool for

   urban children: Results from the Chicago Longitudinal Study. *Minneapolis, MN.*

U.S. Bureau of Labor Statistics. (2002). A century of change: the U.S. labor force, 1950–

   2050. Retrieved from http://www.bls.gov/opub/mlr/2002/05/art2full.pdf

U.S. Census Bureau. (2006/2007) *Income, Poverty, and Health Insurance Coverage in*

   the United States. Retrieved from http://www.census.gov/prod/2007pubs/p60-233.pdf

Valenti, J. E. & Tracey, D. H. (2009). Full-Day, Half-Day, and No Preschool: Effects on

   Urban Children. *Education and Urban Society, 41*(6), 17.

   doi:10.1177/0013124509336060

Waldfogel, J. & Zhai, F. (2008). Effects of public preschool expenditures on the test

   scores of fourth graders: evidence from TIMSS. *Educational Research and*

   Evaluation, 14(1), 9–28.

What Works Clearinghouse. (2012). Research Studies Database. Retrieved from

    http://ies.ed.gov/ncee/wwc/

Whitebook, M. (1989). Who cares? Child care teachers and the quality of care in

    America. Final report, National Child Care Staffing Study. Berkeley, CA: Child Care

    Employee Project.

Williams, J. M., Landry, S. H., Anthony, J. L. & Swank, P. R. (2012).An Empirically-

    based Statewide System for Identifying Quality Pre-Kindergarten Programs.

    Education Policy Analysis Archives, 20(17), 1 – 36.

Wong, V. C., Cook, T. D., Barnett, W. S. & Jung, K. (2008). An effectiveness-based

    evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and*

    Management, 27(1), 122–154. doi:10.1002/jcop.20462

Xiang, Z., Schweinhart, L., Hohmann, C., Smith, C., Storer, E. & Oden, S. (2000). *Points*

    *of Light: Third Year Report of the Michigan School Readiness Evaluation.*High/Scope

    Educational Research Foundation.

Xiang, Z. & Schweinhart, L. J. (2002). *Effects five years later: The Michigan School*

    Readiness Program evaluation through age 10. High/Scope Educational Research

    Foundation.

VITA

Michael J. Haslip

Old Dominion University
Child Study Center
Norfolk, VA 23529-0136
Author Email: mhaslip@gmail.com

EDUCATION

Doctor of Philosophy in Education, Aug. 2013
Early Childhood Education
Old Dominion University, Norfolk, Virginia

Master of Science in Education, May 2008
Early Childhood Education, PreK-3
Old Dominion University, Norfolk, Virginia

Bachelor of Arts in International Development, Oct. 2005
Minor in Political Science
Dalhousie University, Halifax, Nova Scotia

EXPERIENCE

Assistant Professor, 2013 –

Elementary School Teacher , 2010 – 2013

College Instructor, 2008 – 2010

Graduate Assistant, 2007 – 2009

SCHOLARSHIP

Haslip, M.L., Haslip, M. J. (2013). From Malaysia to America: community-based
character education for children and youth. *Childhood Education.* (Forthcoming,
July/Aug. 2013)