UNIVERSITY OF CALIFORNIA,
IRVINE

Promoting School Readiness with Preschool Curricula

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Tutrang Chung Nguyen

Dissertation Committee:
Distinguished Professor Greg J. Duncan, Chair
Founding Dean and Professor Deborah Lowe Vandell
Associate Professor Stephanie M. Reich
Assistant Professor Jade Marcus Jenkins

2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ACKNOWLEDGMENTS

Any acknowledgment for my success in graduate school must begin and end with my advisor, Greg Duncan: begin because I would have never gotten my start as an early childhood researcher if it were not for him; end because he provided unparalleled training that has helped shape my path through UCI and beyond. In many ways, I am reluctant to submit this dissertation because I have loved working with and learning from such a thoughtful and talented scholar and I don't want this time to come to an end. He has inspired me with his ideas, humbled me with his genuine commitment to my learning and development, and imbued me with a passion for contributing to academia. It is an honor and a privilege to be his student, and I hope our collaborations continue as I move on to the next phase of my career.

Jade Jenkins has been a mentor to me in every sense of the word. She taught me essentially everything I know about early childhood education, policy, data analysis, and how to put the three together. It is with her that I first had conversations about these studies and her feedback and ideas that have helped shape them. She has been more gracious and generous than I deserve, and I am most thankful for our collaborations and for her advice and support both personally and professionally. I admire her beyond measure as a successful, hardworking, astute female researcher.

Drew Bailey came to UCI at the beginning of my second year and has been a consistent figure in my graduate experience ever since. Immediately, in ways that only Drew can, he challenged me to think critically, pushed me to communicate clearly, provided guidance when I needed it, and over the years, supported me as I explored my own interests. I am truly grateful for his encouragement, faith in me, and tendency towards a collaborative mentor-mentee relationship.

I also want to thank Stephanie Reich who helped me begin my collaboration with Orange County Head Start and gave me more responsibility than I rightly deserved. I owe much of my applied training to her and the opportunities she has given me. Chapter 3 would not have been possible without her support, enthusiasm, and detailed and discerning feedback.

I must thank Deborah Vandell for not only being on my dissertation committee, but also for being so generous with her time whenever I needed anything. Her excitement and energy for all things research is infectious. She is also an excellent teacher and has profoundly enriched my understanding of human development.

I acknowledge the Institute for Education Sciences and the Interuniversity Consortium for Political and Social Research for making data available, and Orange County Head Start for allowing me to spend time observing in their classrooms. I also appreciate the generous financial support and resources I've received from both the UCI Graduate Division Eugene Cota-Robles Fellowship and the Administration for Children and Families Child Care Scholars Dissertation Grant.

My sincerest gratitude goes to an extraordinary group of friends. I am lucky to have met Joanna Yau Chiou, Melissa Powell Callaghan, and Ryan Lewis during graduate school. From late night working sessions, to celebrating birthdays, marriages, and new jobs, to de-stressing lunches and coffee breaks in the courtyard, they have kept me sane, encouraged, and productive these past five years. Tien Ho and David Liu provided much laughter, welcome distractions and random

adventures, and I will miss their presence in my daily work life. Jordan Lynton and Rose Pier regularly remind me of who I was before graduate school. Whether across the country or right down the street, they have seen me through the failures, the laughter, the tears, and the successes. They have been there before the idea of getting a PhD and have been here through the entire process completing their own degrees in parallel.

Finally, for too many reasons to count, I will never be able to thank my family enough. I am eternally grateful to my sister, who is part of everything I do and aspire to accomplish. Over the past year, like the twenty-two before it, she encouraged me to do my best work, lifted me when I needed it most, fiercely believed in me, supported me unconditionally, and understood me without any need for explanation or elaboration. My parents have always embraced my every endeavor—including the next one I am about to embark on—and provided me with firm shoulders to stand on. They have been incredibly patient and supportive, giving me the space to complete my work and then giving me the strength to continue when I needed encouragement. Anything their daughters accomplish is possible because of all they've sacrificed. This dissertation is dedicated to them.

CURRICULUM VITAE

T U T R A N G C H U N G N G U Y E N

School of Education • University of California, Irvine
2031 Education Building • Irvine, CA • 92697
tutrann@uci.edu • (951) 547-9225
tutrangnguyen.strikingly.com

EDUCATION

2018 **Ph.D. Education**
Specialization in Educational Policy and Social Context
University of California, Irvine

Dissertation: Promoting School Readiness with Preschool Curricula

Committee: Greg J. Duncan (Chair), Jade Marcus Jenkins, Stephanie M. Reich, Deborah Lowe Vandell

2016 **M.A. Education**
University of California, Irvine

2013 **B.A. Mathematics & Liberal Studies**
University of California, Riverside

Honors: Cum Laude, Phi Beta Kappa, University Honors

RESEARCH INTERESTS

Early childhood development, care, and education; child and family policy; program evaluation; using data to improve instruction and systems change

RESEARCH GRANTS

2016-18 U.S. Department of Health and Human Services, Administration for Children and Families, Child Care Research Scholars Dissertation Grant. "Promoting School Readiness with Preschool Curricula: A Mixed-Methods Study." Grant Number 90YE0184, $49,990 (Principal Investigator).

PUBLICATIONS

Nguyen, T., Jenkins, J. M., & Auger, A. (Forthcoming) "Are content-specific curricula differentially effective in Head Start or state pre-k classrooms?" *AERA Open*.

Casasola, T., **Nguyen, T.,** Warschauer, M., & Schenke, K. (2017). Can Flipping the Classroom Work? Evidence from Undergraduate Chemistry. *International Journal of Teaching and Learning in Higher Education, 29*(3), 421-435.

Schenke, K., **Nguyen, T.,** Watts, T. W., Sarama, J., & Clements, D. H. (2017). Differential Effects of the Classroom on African American and Non–African American's Mathematics Achievement. *Journal of Educational Psychology, 109*(6), 794-811.

Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Childhood Research Quarterly, 36*(3), 550-560.

Bailey, D. H., **Nguyen, T.,** Jenkins, J. M., Domina, T., Clements, D. H., & Sarama, J. S. (2016). Fadeout in an early mathematics intervention: Constraining content or preexisting differences? *Developmental Psychology, 52*(9), 1457-1469.

Reimer, L. C., Schenke, K., **Nguyen, T.,** O'Dowd, D. K., Domina, T., & Warschauer, M. (2016). Evaluating Promising Practices in Undergraduate STEM Lecture Courses. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 2*(1), 212-233.

BOOK CHAPTERS

Nguyen, T., Duncan, G.J., & Jenkins, J.M. (In press). Boosting School Readiness with Preschool Curricula. In Reynolds, A.J. & Temple, J.A. (Eds.), *Sustaining Early Childhood Learning Gains: Program, School, and Family Influences*. Cambridge, UK: Cambridge University Press.

Reimer, L., Nili, A., **Nguyen, T.,** Domina, T., & Warschauer, M. (2015). Clickers in the wild: A campus-wide study of student response systems. In Weaver, G. C., Burgess, W. D., Childress, A. L., & Slakey, L (Eds.), *Transforming institutions: Undergraduate STEM education for the 21st century*. West Lafayette, IN: Purdue University Press.

WORK UNDER REVIEW

Nguyen, T., Duncan, R.J., & Bailey, D.H. "Theoretical and Methodological Implications of Associations Between Executive Function and Mathematics in Early Childhood."

Nguyen, T. & Duncan, G.J. "Kindergarten Components of Executive Functions and Third Grade Achievement: A National Study." Second revision under review, *Early Childhood Research Quarterly*.

Jenkins, J.M., Auger, A., **Nguyen, T.,** & Yu, W. "Distinctions without a difference? Preschool curricula and children's development."

WORK IN PROGRESS

Nguyen, T., Li, W., & Duncan, G.J. "Impacts of Targeted and Global Preschool Curricula on Children's School Readiness: A Meta-Analytic Review."

Nguyen, T. "Skill-Specific Preschool Curriculum Interventions: Examining Treatment Effect Heterogeneity with Distributional Methods."

Nguyen, T. "High Scope Curriculum Implementation in Head Start Classrooms: A Qualitative Evaluation."

Nguyen, T., Reich, S.M., Jenkins, J.M., & Abedi, J. "Psychometric Validation of the Desired Results Developmental Profile."

Jenkins, J.M. & **Nguyen, T.** "Keeping Kids in Care: What Makes a Difference in State CCDF Policy?"

HONORS AND AWARDS

2013-18 **Eugene Cota-Robles Fellowship,** University of California, Irvine Graduate Division

2013 **Phi Beta Kappa Graduate Study Award,** Phi Beta Kappa Alumni of Southern California

2013 **Undergraduate Education Research Training Fellowship,** American Educational Research Association

2012 **Rosemary S.J. Schraer Endowed Memorial Scholarship,** University of California, Riverside

2012 **Summer Undergraduate Research Fellowship**, University of California, Irvine Graduate Division Summer Research Program

INVITED PRESENTATIONS

Nguyen, T. "Skill-Specific Preschool Curriculum Interventions: Examining Treatment Effect Heterogeneity with Distributional Methods."

- Curry School of Education, University of Virginia, February 2018
- Neuroscience and Education Lab, Institute of Human Development and Social Change, New York University, December 2017

Nguyen, T. "Promoting Children's School Readiness with Curricula." Presented at the Department of Child, Youth and Family Studies, Institute of Agriculture and Natural Resources, University of Nebraska, Lincoln, January 2018.

SELECTED CONFERENCE PRESENTATIONS

Nguyen, T. "Skill-Specific Preschool Curriculum Interventions: Examining Treatment Effect Heterogeneity with Distributional Methods."

- Society for Research in Child Development Special Topic Meeting on the Use of Secondary and Open Source Data in Developmental Science, Phoenix, AZ, October 2018
- National Research Conference on Early Childhood, Arlington, VA, June 2018

Nguyen, T. "Impacts of Targeted and Global Preschool Curricula on Children's School Readiness: A Meta-Analytic Review." American Educational Research Association Annual Meeting, New York, NY, April 2018.

Nguyen, T. "Distributional Impacts of Academically Targeted Curricula."
- Association for Public Policy Analysis and Management, Chicago, IL, November 2017
- Society for Research in Child Development, Austin, TX, April 2017

Jenkins, J.M. & **Nguyen, T.** "Keeping Kids in Care: What Makes a Difference in State CCDF Policy?"
- Society for Research in Child Development, Austin, TX, April 2017
- Association for Public Policy Analysis and Management, Washington, DC, November 2016

Nguyen, T. "Assessing Children's Growth in School Readiness by Preschool Curricula Type." Association for Public Policy Analysis and Management, Washington, DC, November 2016.

Nguyen, T. "Contributions of Executive Functions to First Grade Achievement: A National Study." American Educational Research Association Annual Meeting, Washington, DC, April 2016.

Jenkins, J.M., Auger, A., Yu, W., & **Nguyen, T.** "Distinctions without a difference? Preschool curricula and children's development." Association for Public Policy Analysis and Management, Miami, FL, November 2015.

Nguyen, T., Jenkins, J. M., Auger, A., & Domina, T. "Comparing the Effectiveness of Targeted Curricula in Head Start and Public Pre-K Classrooms."
- Association for Public Policy Analysis and Management, Miami, FL, November 2015
- Society for Research on Educational Effectiveness, Washington, DC, March 2015

Nguyen, T., Watts, T.W., & Duncan, G.J. "Which specific preschool mathematics skills predict later mathematics achievement?" Society for Research on Educational Effectiveness, Washington, DC, March 2015.

RESEARCH POSITIONS

2016-17 **Graduate Student Researcher** (with Stephanie M. Reich & Jade Marcus Jenkins)
"Making Orange County Head Start More Effective: A Researcher-Practitioner Partnership to Improve School Readiness for Low-Income Children"

2016 **Graduate Student Researcher (**with Jade Marcus Jenkins)
"Child Care Subsidies and State-Level Policies"

2014-15 **Graduate Student Researcher** (with Greg J. Duncan)
"Early Childhood Mathematics Skills"

2013-14 **Graduate Student Researcher** (with Mark Warschauer & Thurston Domina)
"Documenting Instructional Practices in STEM Lecture Courses"

2012 **Visiting Undergraduate Student Researcher** (mentored by Thurston Domina)
"State-Level Education Standards for School-Based Substance Use Prevention Programs"

TEACHING EXPERIENCE

Spring 2017 **Education 265: Applied Regression,** Teaching Assistant & Stata Lab Instructor
Teaching Assistant with dedicated lab section for PhD students in Education, Sociology, Psychology & Social Behavior, and Nursing.

Fall 2016 **Education 10: Education Research Design,** Teaching Assistant
Teaching Assistant with three dedicated discussion sections for undergraduate students in the Education Sciences major.

Summer 2015, 2016 **Education 30: 21st Century Literacies,** Teaching Assistant
Teaching Assistant for an online undergraduate elective course in the Education Sciences major.

PROFESSIONAL SERVICE

Journal Referee

Developmental Psychology, Early Childhood Research Quarterly, Journal of Children and Poverty

Grant Referee

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

UNIVERSITY SERVICE

2016-18 **Advisory Board Member,** Graduate Resource Center, University of California, Irvine

2016-17 **Ph.D. Admissions Committee,** University of California, Irvine School of Education

2016-17 **Co-chair, Diverse Educational Community and Doctoral Experience**, University of California, Irvine School of Education

2015-17 **Elected Cohort Representative**, Association of Doctoral Students in Education, University of California, Irvine School of Education

2014-16 **Graduate Student Mentor,** Diverse Educational Community and Doctoral Experience Competitive Edge Program & Summer Undergraduate Research Fellowship Program**,** University of California, Irvine Graduate Division

SKILLS

Analytic Techniques

Experimental and quasi-experimental methods, meta-analysis, descriptive statistics, multiple regression, structural equation modeling, hierarchical models, classroom observations and evaluation

Software

Stata, R, SPSS, Atlas.ti, Comprehensive Meta-Analysis (CMA), Tableau

Languages

Fluent Vietnamese

PROFESSIONAL TRAINING

2017 **Introduction to the High Scope Curriculum for Preschool Teachers,** High Scope Educational Research Foundation, Ypsilanti, MI

2017 **ICPSR Summer Program Short Course on Qualitative Research Methods,** University of North Carolina, Chapel Hill Odum Institute

2017 **Classroom Assessment Scoring System (CLASS),** Observer Reliability Training for Preschool, Santa Monica, CA

PROFESSIONAL MEMBERSHIPS

American Educational Research Association
Association for Public Policy Analysis and Management
Child Care and Early Education Policy Research Consortium
Society for Research in Child Development
Society for Research on Educational Effectiveness

ABSTRACT OF THE DISSERTATION

Promoting School Readiness with Preschool Curricula

By

Tutrang Chung Nguyen

Doctor of Philosophy in Education

University of California, Irvine, 2018

Distinguished Professor Greg J. Duncan, Chair

Preschool curricula are a much-researched topic in early childhood education and a renewed emphasis has been placed on understanding the effectiveness of curricula as the number of children enrolled in public preschool in the U.S. is increasing. Key to understanding how curricula are more or less effective in promoting classroom quality and children's school readiness is knowing the differential relations between various packages that are used in public preschool classrooms, the extent to which these curricula are being implemented faithfully, and for whom they are most effective.

This dissertation, comprised of three studies, comprehensively examines the different ways in which various curricula promote children's learning and development so that they are school-ready at the end of the preschool year. The first and second studies focus on whole-child curricula. For study one, I examine associations of various whole-child curricula packages on classroom quality and children's school readiness outcomes using data from the Head Start Family and Child Experiences Survey, 2009 Cohort. Results suggest that the whole-child curricular packages under examination have no associations with improved classroom quality or children's school readiness skills, including academic, socio-emotional, and executive function outcomes. In study two, I conduct open-ended observations of curriculum implementation

fidelity across four classrooms in local Orange County, California Head Start centers to understand what a whole-child classroom looks like in practice. I find that teachers are able to implement easier, explicit components of the curriculum throughout the day as these activities are built into the daily classroom routine. However, I observe that teachers struggle with scaffolding children's learning during these activities, which takes considerable skill on their part. Finally, in the third study I use quasi-experimental methods to explicitly test five different hypotheses of treatment effect heterogeneity in academic skill-specific curriculum interventions based on educational, developmental, and economic theory. Findings suggest that skill-specific curricula have differential effects at the top and bottom of the distribution of children's literacy and language outcomes.

Together, the three studies provide new evidence that furthers the field's understanding of curriculum as an important instructional feature in public preschools, both nationally and locally, and has implications for promoting the development of linguistically and culturally diverse low-income and disadvantaged children. These results provide policy-relevant information for facilitating the most efficient use of early learning funding, making decisions about how best to target specific curricular programs, and suggesting ways to improve the design and implementation of programs for high-quality preschool.

CHAPTER ONE

Introduction to the Dissertation

High-quality preschool programs can improve children's school readiness and future academic success (Karoly, Kilburn, & Cannon, 2005; Magnuson, Ruhm, & Waldfogel, 2007), particularly for children from low-income backgrounds who are more likely to be at risk of early school failure (Barnett, 2011; Ramey & Ramey, 2006; Schweinhart, 2006). Researchers, educators, and policymakers have pinpointed preschool programs' use of curriculum-based interventions as a means to increase school readiness and narrow the achievement gap. From a theoretical perspective, a strong curriculum can improve preschool classroom quality by setting goals for the knowledge and skills that children should gain in an educational setting and supporting educators' plans for providing the day-to-day learning experiences to cultivate those skills through daily lesson plans, materials and other pedagogical tools (Goffin & Wilson, 1994; Ritchie & Willer, 2008; Klein & Knitzer, 2006; National Association for the Education of Young Children (NAEYC) & National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE), 2003).

The majority of publicly-funded preschool programs, including Head Start and state preschool, require use of a "research-based curriculum." Two broad categories of curricula include: (1) 'whole-child' curricula, which aims to provide enriching experiences across multiple domains of children's development (e.g., health, social-emotional, academic), and (2) more targeted, skill-specific curricula. In the case of whole-child curricula, there is limited empirical evidence of how it relates to classroom and child outcomes. Most publishers claim their packages are research-based, yet very few describe the research on which the claim is based or

how they are explicitly linked to improved classroom quality or children's school readiness (Clements, 2007).

An important piece of the puzzle to this question of overall effectiveness is how curricula are implemented in the classroom. To date, curriculum implementation in preschool has largely been captured with little detail about the idiosyncratic aspects of classroom processes. In the case of skill-specific curricula, there has been a growing evidence base of the effectiveness of 'skill-specific' curricula on directly targeted school readiness domains (Bierman et al., 2008; Clements & Sarama, 2008; Diamond, Barnett, Thomas, & Munro; Morris et al., 2014). However, children likely enter preschool with differing levels of skills, and it is not entirely clear for whom skill-specific curricula may be more or less effective. Addressing these issues surrounding curricula will help guide policymakers and program directors on how to facilitate the most efficient use of early learning funding, make informed decisions about how best to target specific curricular programs, and improve the design and implementation of programs for high-quality preschool.

The goal of this three-study dissertation is to comprehensively examine the different ways in which various curricula promote children's learning and development so that they are school-ready at the end of the preschool year. Through the use of descriptive, qualitative, and quasi-experimental methods, I ask: (1) Are widely used preschool curricula packages differentially related to classroom quality and children's school readiness skills? (2) How do teachers implement a whole-child curriculum across different classrooms? (3) What is the effect of skill-specific curricula on the distribution of children's achievement?

In the remainder of this chapter, I briefly summarize three empirical studies in which I address many of the research gaps identified above and consider the significance of this dissertation as a whole. In the subsequent three chapters, I fully describe each of these studies. I

conclude this dissertation in Chapter 5 by discussing the substantive, methodological, and policy implications of my findings and directions for future research.

Study 1

My first study, co-authored with Jade Jenkins and Ana Auger, examines the associations of various whole-child curricula on classroom quality and children's school readiness outcomes using data from the Head Start Family and Child Experiences Survey, 2009 Cohort. Despite the popularity of whole-child curricula in preschool programs, there is little to no empirical support of its effectiveness in the literature. The evidence that does exist included a small sample of children from the 1960s with counterfactual conditions that no longer apply to ECE today (i.e., no preschool at all; Belfield, Nores, Barnett, & Schweinhart, 2006; Preschool Curriculum Evaluation Research (PCER), 2008; Schweinhart, 2005). In my study, I use Head Start grantee fixed effects models to reduce curricula selection bias and find that the whole-child curricular packages under examination have no associations with improved classroom quality or children's school readiness skills, including academic, socio-emotional, and executive function outcomes. These results call into question whether the current curricular investments in early childhood policy produce any benefits for children's development.

Study 2

The second study of this dissertation qualitatively evaluates whether and how teachers implement a whole-child curriculum with fidelity as the developers had intended. I conduct open-ended observations of curriculum fidelity across four classrooms to help explain implementation as it is happening in authentic preschool settings. These observations take place in Head Start centers with a local partner, Orange County Head Start, which provide services to nearly 4,000 low-income preschoolers and families throughout the county. I observe the ways in

which teachers implement the curriculum through their daily routine, the arrangement of the learning environment across classrooms, discrete instructional activities for large- and small-group time, and child-managed activities such as centers or free play time. Informal conversations with each teacher enriched the observational data, and lesson plans and classroom artifacts are also collected to triangulate findings.

I find that teachers implement some of the High Scope curriculum goals, but not others. Specifically, teachers implement the "Plan-Do-Review" model of High Scope easily because it is built into their daily routine. Teachers also focus their lesson plans on children's interests, and true to the whole-child philosophy, they incorporate all school readiness learning domains for children with the help of curriculum content indicators. However, teachers struggle with scaffolding children's learning throughout the day, which takes considerable skill on their part. Results suggest that preschool centers may do well to reconsider what teachers supports, such as coaching or additional professional development, are needed to help them provide this type of "light-touch" instruction that promotes the cumulative development of school readiness skills for children over the course of the preschool year.

Study 3

My final study uses data from Preschool Curriculum Evaluation Research (PCER) Initiative to examine the effect of skill-specific literacy curricula on the distribution of children's achievement in preschool. I explicitly test five different hypotheses of heterogeneous treatment effects based on educational, developmental, and economic theory. I pool the PCER data together from all grantees that implemented: 1) a literacy curriculum where the comparison condition was Creative Curriculum or High Scope—the two most commonly used curricular packages in early childhood classrooms; or 2) a literacy curriculum where the comparison

condition was a locally-developed curriculum. I find that skill-specific literacy curricula have differential effects at various points across the distribution of children's literacy and language skills depending on the developmental outcome. Specifically, I observe a *compensatory* effect on the Woodcock-Johnson Letter Word measure in that children at the bottom of the distribution benefited most from the literacy curricula treatment. I find a *skills beget skills* effect for children at the top of the distribution on Woodcock-Johnson Spelling. For the domain of receptive vocabulary, as measured by the PPVT, I find that results were consistent with the *no effects* hypothesis. One of the main contributions of this study is to familiarize researchers with distributional analytic techniques as a way to understand treatment effect heterogeneity in early childhood interventions.

Significance and Implications

Empirically supported preschool curricula can ensure that children are provided with opportunities to learn by guiding the nature of instruction and the availability of materials and activities in the classroom. Given current efforts at the federal and state levels to expand preschool programs and increase quality, and the dollars invested each year on curricula for these programs, understanding whether different packages are more or less effective in improving classroom processes and promoting school readiness skills and for whom is essential. Between purchasing the teacher's manual, classroom materials, tools for monitoring and assessment, training teachers, and providing coaching or professional development, the cost per classroom can be a large investment. Further, there is a considerable need for more research on factors that enhance or relate to adequate implementation of curricular programs and lead to effective classroom practices and child outcomes. This is particularly important for preschool programs where there is a growing emphasis on the implementation of "research-based" curricula. Thus,

understanding how different types of curricula vary in effectiveness and implementation in promoting classroom and child outcomes for linguistically and culturally diverse low-income and disadvantaged children is a national and state-level early childhood policy concern.

I aim to inform early childhood education policies by directly addressing the lack of research on curricular influences on children's outcomes, and ultimately, the gap in achievement between high- and low-income children that preschool programs aim to address. The results together have implications for early education policy decision-making and program administration and address a particular feature of early childhood education that is important in supporting children's learning and development. Taken together, the three studies provide useful evidence in what should surely be a more extensive and ongoing research agenda.

References

Barnett, W. S. (2011). Effectiveness of educational intervention. *Science, 333*(6045), 975-978. doi: 1.1126/science.1204534

Belfield, C. R., Nores, M., Barnett, S., & Schweinhart, L. (2006). The High/Scope Perry preschool program cost–benefit analysis using data from the age-40 follow up. *Journal of Human Resources*, *41*(1), 162-190. doi: 10.3368/jhr.XLI.1.162

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., . . . Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 79*(6), 1802-1817. doi: 10.1111/j.1467-8624.2008.01227.x

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula". *Journal for Research in Mathematics Education,* 35-70. doi: 10.2307/30034927

Clements, D., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*(2), 443-494. doi: 10.3102/0002831207312908

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387-1388. doi: 1.1126/science.1151148

Goffin, S. G., & Wilson, C. (1994). *Curriculum models and early childhood education: Appraising the relationship*. New York, NY: Merrill.

Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2005). *Early childhood interventions: Proven results, future promise*. Santa Monica, CA: RAND Corporation.

Klein, L., & Knitzer, J. (2006). *Effective preschool curricula and teaching strategies*. New York, NY: National Center for Children in Poverty.

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

Morris, P. A., Mattera, S. K., Castells, N., Bangser, M., Bierman, K. L., & Raver, C. C. (2014). *Impact findings from the Head Start CARES Demonstration: National evaluation of three approaches to improving preschoolers' social and emotional competence.* Washington, DC: Office of Planning Research and Evaluation, Administration for Children and Families.

National Association for the Education of Young Children (NAEYC) and National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). (2003). *Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age 8. Joint position statement.* Washington, DC: National Association for the Education of Young Children.

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness*. (NCER 2008-2009). Washington, DC: National Center for Education Research, Institute for Education Sciences, U.S. Department of Education.

Ramey, C. T., & Ramey, S. L. (2006). Early learning and school readiness: Can early intervention make a difference? In N. F. Watt, C. Ayoub, R. H. Bradley, J. E. Puma & W.A. LeBoeuf (Eds.), *The crisis in youth mental health* (pp. 291–318). Westport, CT: Praeger.

Ritchie, S., & Willer, B. (2008). *Curriculum: A guide to the NAEYC early childhood program standard and related accreditation criteria*. Washington, DC: National Association for the Education of Young Children (NAEYC).

Schweinhart, L. J. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40.* Ypsilanti, MI: High/Scope Educational Research Foundation.

Schweinhart, L. J. (2006). The High/Scope approach: Evidence that participatory learning in early childhood contributes to human development. In N. F. Watt, C. Ayoub, R. H. Bradley, J. E. Puma & W.A. LeBoeuf (Eds.), *The crisis in youth mental health* (pp. 207-227). Westport, CT: Praeger.

CHAPTER TWO

Study 1: Whole-Child Curricula and their Associations with Classroom Quality and Children's School Readiness

Curricula are key features of publicly funded early childhood education programs—Head Start and state pre-kindergarten—and monitoring initiatives such as Quality Rating and Improvement Systems (QRIS). Head Start programs are mandated to use curricula that are centered on educating the "whole-child", or in other words take a more 'global' approach to child development. These curricula typically take a constructivist approach to learning, emphasizing child-centered active learning that is promoted through intentional arrangement of the classroom environment (DeVries & Kohlberg, 1987; Piaget, 1976; Weikart & Schweinhart, 1987). Rather than explicitly targeting academic domains such as early math skills, whole-child curricula aim to support children's learning and development by encouraging them to interact independently with materials and other children in the classroom. Whole-child curricula focus on improving a broad range of children's school readiness skills, including cognitive, physical, emotional, and pre-academic outcomes. This type of curricula reflects the standards for early childhood education agreed upon by the National Association for the Education of Young Children—the leading professional and accrediting organization for early educators (Copple & Bredekamp, 2009). Despite the support and adoption of whole-child curricula in the majority of today's preschool classrooms, we know very little about how it relates to classroom quality and promoting children's school readiness.

In the current study, I seek to understand whether commonly used preschool curricula differentially promote classroom quality and children's school readiness skills using data from a large sample of low-income three- and four-year old children attending Head Start programs

across the U.S. I estimate a series of models to examine different curricular packages and their relations with classroom quality and child school readiness outcomes. The results of this study will help inform practice and policy related to curricular choices in Head Start programs who are mandated to use research-based whole-child curricula.

Background

The foundation of whole-child curricula is a pedagogical philosophy that is playful in delivery. In a classroom implementing a whole-child curriculum, the teacher's role is to support and encourage children's play and exploration which can require considerable skill on the part of the teacher. Children are free to engage with different components of the classroom environment, and the teacher's task is to support or "scaffold" learning with just the right amount of input. While carefully observing what children say and do, teachers are expected to develop an understanding of how the children think and then further learning experiences they need. Teacher input is meant to promote learning in a way that is not so directive that children begin to lose interest because of the instruction. Additionally, the sequence of inputs provided by the teacher should promote cumulative development of academic or social-emotional skills over the course of the preschool year, which is possibly the most difficult goal of all to accomplish.

The vast majority of preschool curricula are created by educational researchers and practitioners and then sold to practitioners by publishers. Whole-child curricula most widely used by preschool classrooms in Head Start programs are Creative Curriculum and High Scope, and packages such as Scholastic and High Reach curricula are other common alternatives (Clifford et al., 2005; Hulsey et al., 2011; Phillips, Gormley, & Lowenstein, 2009). These curricular packages all share a broad focus on developing children's academic and social-emotional skills and promoting health and nutrition. Moreover, these curricula also allow

children a large proportion of free-choice time (Fuligni, Howes, Huang, Hong, & Lara-Cinisomo, 2012).

Federally and state-sponsored QRIS incorporate curriculum into their rankings and consider the use of a developmentally appropriate, research-based curriculum an indication of program quality (Auger, Karoly, & Schwartz, 2015). Though most publishers claim their packages are research-based, very few describe the research on which the claim is based or how they are explicitly linked to improved classroom processes or children's school readiness (Clements, 2007). Apart from the Perry Preschool Study, little empirical support exists for High Scope (Belfield, Nores, Barnett, & Schweinhart, 2006; PCER, 2008; Schweinhart, 2005). Results from Perry indicate that children who attended the program and used its High Scope curriculum tended to benefit more in the short- and long-term compared with children who did not attend the program. However, this study included a small sample of children from the 1960s with counterfactual conditions that no longer apply to ECE today (i.e., children who did not attend center-based preschool and very low levels of maternal schooling). Further, no empirical support exists for Creative Curriculum, and it has not demonstrated effectiveness based on rigorous What Works Clearinghouse standards (U.S. Department of Education, 2013), despite its popularity in preschool classrooms (Hulsey et al., 2011).

The *Preschool Curriculum Consumer Report*, released by the National Center on Quality Teaching and Learning of the Office of Head Start (2014), reviewed the most commonly used curricula in Head Start programs nationwide and provided ratings for each on a set of different criteria. This report was the first of its kind to give providers guidance about the selection of the most appropriate curricula or combination of curricula for their programs. One criterion is whether the curriculum is evidence-based. Of the 14 curricula reviewed in the report, seven had

"no evidence", five had "minimal evidence", one had "some evidence", and only one (Opening the World of Learning) was rated to have "solid, high-quality evidence" with demonstrated effects on child outcomes. Creative Curriculum and High Scope were both rated as having no evidence of effects on child outcomes.

Little support for whole-child curricula that are being implemented in the majority, if not all, of Head Start programs across the U.S. is a great concern. Given the millions of federal and state dollars spent each year on curricula for public preschool programs operating at scale and the mixed findings to date on whole-child curricula, more evidence is needed on the effect of these curricular programs in improving classroom processes and preparing children for kindergarten across a host of domain-specific (e.g., math and literacy) and domain-general (e.g., socio-emotional and executive function) school readiness skills.

Method

Data

The current study uses data from the Head Start Family and Child Experiences Survey, 2009 Cohort (FACES 2009; U.S. DHHS, 2009-2013). Head Start FACES was launched in 1997 as a periodic, longitudinal study of program performance. Successive nationally representative samples of Head Start children, their families, classrooms, and programs were collected in 2000, 2003, 2006, and 2009 to provide descriptive information on the population served, staff qualifications, Head Start classroom practices and quality measures, and child and family outcomes. These data provide a rich source of ongoing information on the children and families served by Head Start, the programs themselves, and on the staff. The FACES sampling design included a four-stage sampling process to select a representative group of Head Start (1) grantees (programs); (2) centers; (3) classrooms; and (4) newly enrolled children. Sampling at the first

three stages was done with probability proportional to size. In total, the FACES 2009 sample included 60 grantees, 129 centers, 486 classrooms, and 3,349 children. Data collection took place from fall 2009, when children entered Head Start, until the end of kindergarten, either in spring 2011 or 2012. This study uses data from FACES 2009, the last year in which FACES collected longitudinal data.[1] I analyze data from fall 2009 (collected over a nine-week period that began in late September and concluded in mid-November) to spring 2010 (collected between mid-April and mid-June).

Measures

Curricula. The FACES 2009 research team used teacher interview instruments to collect data on a variety of curricula packages of interest to this study. More specifically, although teachers were allowed to select multiple published curricula used in their classrooms, the FACES study also asked teachers what the primary curriculum was that they used in class, which I use as the key independent variable. All curriculum packages reported in FACES 2009 are whole-child. These include Creative Curriculum, High Scope, Scholastic, and High Reach. The "other" curricula category includes less commonly used packages such as Let's Begin with the Letter People, Montessori, and Bank Street.

Classroom quality. Quality of the classroom environment was measured with several instruments in FACES in the spring of 2010. The first is the *Early Childhood Environment Rating Scale-Revised* (ECERS-R; Harms, Clifford, & Cryer, 1998) is a widely used observer-rated measure of global classroom quality, specifically designed for use in classrooms serving children between 2.5 and 5 years of age. Scores on the ECERS-R range from 1-7 with 1

[1]Successive samples of Head Start children, their families, classrooms, and programs have been collected in 2014 and data collection is ongoing up until 2022. However, in 2011 the Administration for Children and Families redesigned the study. FACES is no longer longitudinal, but now studies all enrolled children within a program year.

indicating “inadequate” quality, 3 indicating “minimal” quality, 5 indicating “good” quality, and 7 indicating “excellent” quality. The scale’s authors report a total scale internal consistency of .92. The second measure is the *Classroom Assessment Scoring System* (CLASS; Pianta, La Paro, & Hamre, 2008), an observer-rated assessment of teacher-child interactions in terms of emotional support (climate, teacher sensitivity, regard), classroom organization (behavior management, productivity, instructional learning formats), and instructional support (concept development, feedback quality, language modeling). The Cronbach’s α are .88 for Classroom Organization, .90 for Emotional Support, and .93 for Instructional Support. ECERS and CLASS scores were collected in the spring of 2010.

Language and literacy skills. This study draws upon three commonly used language and literacy outcomes. Children were assessed at the fall and spring of the Head Start year. The first measure is the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997), which assesses children’s vocabulary knowledge. It takes approximately 5 to 10 minutes to complete and is administered by a trained researcher. The measure requires the child to point to the picture that represents the word spoken to them by the researchers. Words increase in difficulty and scores are standardized according to the child’s age. The measure is nationally normed, with a mean of 100 and a standard deviation of 15.

The second and third language and literacy measures, Letter Word and Spelling, come from the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001). The Letter Word subtest is similar to the PPVT in that it asks children to identify the letter or word spoken to them, and the test gradually increases in difficulty to require the child to read words out of context. The Spelling subtest requires children to write and spell words presented to them. Both assessments from the WJ-III were administered by trained researchers and each took

approximately 10 minutes to administer. Like the PPVT, scores are standardized by the age of the child and nationally normed to have a mean of 100 and a standard deviation of 15. For the purposes of the analyses, the assessments are standardized for the sample to have a mean of 0 and a standard deviation of 1 for ease of interpretation.

Math skills. To measure children's mathematics knowledge, children were assessed at the fall and spring of the preschool year with the Applied Problems subtest of the WJ-III. This measure requires children to solve increasingly difficult math problems. This instrument also assesses basic mathematics skills such as number recognition. This subtest is standardized for the child's age and nationally normed to have a mean of 100 and a standard deviation of 15. The assessment takes approximately 10 minutes to administer. For the analyses, this mathematics measure is standardized for the sample to have a mean of 0 and a standard deviation of 1.

Social-emotional behaviors. Problem behaviors and social skills were measured in the FACES studies using items from an abbreviated adaptation of the Personal Maturity Scale (Alexander, Entwisle, Blyth, & McAdoo, 1988), the Child Behavior Checklist for Preschool-Aged Children, Teacher Report (Achenbach, Edelbrock, & Howell, 1987), the Behavior Problems Index (Zill, 1990), and the Social Skills Rating System (Gresham & Elliott, 1990). Teachers reported children's social-emotional behaviors in the fall and spring of preschool. For the analyses, the social skills and problem behaviors measures are standardized within the sample to have a mean of 0 and a standard deviation of 1.

Executive function. Children's executive function was directly assessed with a pencil tapping task at the fall and spring of preschool. This task is an adaptation of a peg-tapping task (Blair, 2002; Diamond & Taylor, 1996) and requires the child to do the opposite of what the assessor says. For example, the child is asked to tap one time when the assessor says to tap two

times and tap two times when the assessor says to tap one time. FACES 2009 is the first FACES cohort to use this task. Because this task requires working memory and attention to auditory and visual stimuli, it was only administered to children ages 4 and older. I standardized this measure within the sample to have a mean of 0 and a standard deviation of 1.

Covariates. I include a host of covariates in all analyses that were collected at baseline during the fall of the preschool year. At baseline the primary caregiver reported on child, personal, and family demographics and background characteristics. Information on classrooms and teachers were drawn from interviews or surveys that teachers and center directors completed about their classrooms and themselves. These characteristics included teachers' education, race, and years of experience. Child-level characteristics included gender (1 = female), race (white as the reference category, black, Asian, Hispanic, and other), and age in months. Maternal or primary caregiver and family characteristic included education level, age, and annual household income in thousands of dollars. Also included in the analyses are children's fall school readiness measures, along with classroom measures where appropriate.

Analysis Plan

I conducted four sets of analyses focused on classroom and child outcomes. First, I compare Creative Curriculum with all other whole-child curricula to examine whether the most commonly used curriculum package in Head Start centers is predictive of children's outcomes compared with any other alternative curriculum package. All other curriculum packages are pooled together. The general form of this model is as follows:

$$(1)\ \text{Outcome}_{ic} = b_0 + ß_1(\text{Creative Curriculum}_{ic}) + ß_2(\text{Covariates}_{ic}) + \varepsilon_{ic}$$

where Outcome_{ic} represents the classroom or child outcomes observed for child *i* in classroom *c* at the end of the preschool year; Covariates_{ic} is a vector of classroom, child, and family control

variables for child *i*; and ε_{ic} represents the remaining sources of variation in children's school readiness from unaccounted factors. $ß_1$ is the coefficient of interest, representing the association between a classroom using Creative Curriculum or not, and the classroom or child outcomes, indexed by classroom *c*.

For the second set of analyses, I add in grantee fixed effects, which are used in the analyses to remove any context-specific and time-invariant observable or unobservable characteristics that may influence both the choice of curriculum and children's outcomes. That is, these models will control for differences in a context shared by children, reducing the possibility of omitted variables bias. A check of the data indicates there is sufficient variation in different curricula used across grantees for a fixed effects model. Twenty-eight (47%) out of the 60 grantees had variation across classrooms in curricular package. The form of this model is shown below:

$$(2)\ \text{Outcome}_{icz} = b_0 + ß_1(\text{Creative Curriculum}_{ic}) + ß_2(\text{Covariates}_{ic}) + \delta_z + \varepsilon_{ic}$$

where δ_z is a vector of indicators for each of the Head Start grantees *z*, defined as programs or delegate agencies with the funding to provide direct services, included in the study, and all other terms are the same as before.

In the third and fourth set of analyses, I replace $ß_1$ in Equations (1) and (2) with a vector of curriculum indicator variables which vary by classroom. These coefficients are the key estimates of interest since they represent the differential associations between each preschool curriculum and the outcomes of interest relative to the reference category. Creative Curriculum was the most common package used in the data, and thus serves as the reference category. The other curriculum indicator variables include: High Scope, High Reach, Scholastic, and other published curricula (listed in Table 1's footnote).

All sets of analyses use Ordinary Least Squares regression and adjust for the clustered sample designs at the grantee-level using Huber-White standard errors and fixed effects where appropriate. Missing data were handled using dummy variables (1 = missing). Missing values on the covariates were then set to zero and the missing dummies were then entered into the regressions.

Classroom outcome analyses. The sample for the classroom outcomes analyses included children in classrooms for whom at least one of the classroom observational measures (ECERS or CLASS) was available. The classroom and teacher covariates included teachers' education, race, years of experience, and classroom-level aggregates of children's race, gender, and parental education. These classroom-level aggregates were included to account for student composition within a classroom during the observational assessments. Because ECERS and CLASS scores were collected in the spring of 2010, we did not have the corresponding fall of 2009 score to include as a covariate in our models.

Child outcome analyses. The sample for the child outcomes analyses consisted of children who had at least one school readiness outcome at the end of preschool. Covariates included gender of child, race of child, mother or primary caregiver educational level and age, and family income. I also included children's baseline outcome assessments from the fall of the preschool year as covariates.

Results

Descriptive Analyses

Prior to conducting substantive analyses, descriptive statistics were computed for the sample by curricular package. Table 1 presents the descriptive statistics and ANOVAs on the covariates used in the analyses, and the outcomes, baseline academic, and social-emotional

scores for the sample by curricular package. Few differences in children's baseline test scores emerged across the different curricula. There were significant differences across curricula packages in overall classroom quality based on the ECERS and CLASS scales. Other significant differences emerged between curricular packages in each dataset, but without a clear rank ordering of packages in terms of superior quality (e.g., High Scope did not have consistently higher scores on the CLASS dimensions than Creative Curriculum). Additionally, there were some significant differences in teacher education level and center characteristics across the different curricular packages.

[Insert Table 2.1]

Substantive Analyses

Results of the substantive analyses for the classroom quality outcomes are displayed in Tables 2.2 and 2.3. Results for the child outcomes are presented in Tables 2.4 and 2.5. All dependent variables are standardized to have a mean of zero and standard deviation (SD) of one so that the coefficients can be understood as effect sizes. To orient the reader to how the results are presented, I discuss the first column of each respective outcome (Model 1), followed by the second column which includes grantee-level fixed effects (Model 2). Model 1 coefficients on the curriculum indicators can be interpreted as the difference between a classroom using a particular curriculum relative to Creative Curriculum on the classroom quality or school readiness outcome of interest. With the addition of grantee-level fixed effects, Model 2 coefficients on the curriculum indicators can be interpreted as the difference in classroom or child outcomes between a classroom using a particular curriculum and Creative Curriculum classrooms operated by the same grantee.

Classroom quality. Table 2.2 presents regression results for classroom quality outcomes for classrooms using Creative Curriculum compared with classrooms using any other published curriculum. Model 1 indicates a marginally significant negative association ($p < .10$) between classrooms using Creative Curriculum and classrooms using any other published curriculum on the ECERS but no association on the CLASS. Model 2 shows that there were no statistically significant associations on the ECERS or the CLASS between classrooms using Creative Curriculum compared with classrooms operated by that same program using all other published curricula.

[Insert Table 2.2]

Table 2.3 shows regression results for the associations between the various curricula packages and classroom quality. Model 1 indicates that classrooms implementing the High Reach curriculum had significantly lower associations with classroom quality as measured by the ECERS (-0.34 SD, $p < .05$). Classrooms using Scholastic and other published curricula had marginally significantly lower ECERS scores (-.34 and -.32 SD, respectively, $p < .10$). There were no other significant associations with the ECERS. There were also no significant associations between the various curricula and the CLASS. Model 2 includes the grantee-level fixed effects and indicates that classrooms using High Reach scored significantly lower on the ECERS than classrooms using Creative Curriculum operated by that same Head Start program (-.31 SD, $p < .05$). There were no other significant associations with the ECERS or CLASS for Model 2. I conducted F-tests of the four included curricula indicators to test for the joint significance of the curriculum packages. The p-values of these tests are shown in the bottom of Table 2.3. The coefficients for the different curricula packages were significantly different from one another for the ECERS, but not significantly different from one another for the CLASS.

High Reach classrooms were rates as much worse on the global scale of classroom quality (ECERS) than any of the other classrooms implementing different curricula.

[Insert Table 2.3]

Child outcomes. Table 2.4 presents results for models examining differences in school readiness outcomes between children in classrooms using Creative Curriculum compared with children in classrooms using all other curricula. Across Models 1 and 2, no significant differences in children's outcomes emerged, suggesting that children in classrooms implementing the most widely used curriculum package across Head Start centers are doing no better than classrooms implementing any other published curricula.

[Insert Table 2.4]

Models examining differences in children's gains in school readiness outcomes by curricular package are shown in Table 2.5. Model 1 results show that children in classrooms using High Reach scored significantly lower than children in classrooms using Creative Curriculum on the PPVT (-.30, $p < .05$) and WJ Applied Problems (-.25, $p < .01$). There were no other statistically significant associations between the other curricular packages and children's outcomes. Turning to Model 2 of each respective outcome that includes grantee-level fixed effects, the results indicate that children in Head Start classrooms using the High Reach Curriculum scored significantly worse than children in other classrooms operated by that same program using the Creative Curriculum on the PPVT (-.33 SD). For the WJ Applied Problems scores, children scored significantly lower in classrooms using the High Reach Curriculum when compared with Creative Curriculum (-.18 SD).

[Insert Table 2.5]

I conducted F-tests of the four included curricula indicators to test for the joint significance of the curriculum packages. The p-values of these tests are shown in the bottom of Table 2.5. For the PPVT and the WJ Applied Problems the coefficients for the different curricula packages were significantly different from one another. However, for the WJ Letter Word, WJ Spelling, social-emotional behaviors, and executive function, the coefficients of the different curricula packages were not significantly different from one another.

Discussion

The goal of this study was to understand the relative effects of policy-mandated whole-child curricula on classroom quality and children's school readiness skills. Given the call for scientifically validated curricula in large-scale public preschool programs such as Head Start, this analysis contributes to a body of literature examining how different curricula influence classroom processes and children's outcomes prior to kindergarten entry. Taken together, the findings from these two studies indicate that there are few distinguishing characteristics of whole-child curricula most commonly used in today's preschool programs (e.g., Creative Curriculum and High Scope) and inconsistent evidence of differential associations between various curricular packages and children's school readiness. One general pattern observed from the results is that High Reach classrooms had significantly lower quality scores and children scored lower on both academic and social skills outcomes than children in Creative Curriculum classrooms. These results are in line with the What Works Clearinghouse (2013) rating of "No Evidence" for Creative Curriculum, and there is little support for the other curricula included in this study.

It is important to note here in the interpretation of the results that all of the curricula could have been highly effective, but just not differentially effective. If Creative Curriculum or

whole-child curricula in general were compared with no curricula or locally-developed curricula in which teachers or programs develop materials on their own, it could be argued that they may be predictive of classroom quality or children's outcomes. This was not possible to examine with the current study data since all Head Start classrooms are mandated to use a whole-child curriculum and thus there was no category of locally-developed curricula, but there is some recent evidence in the literature suggesting that children exposed to whole-child curricula do not outperform children in classrooms where curricula are developed locally (Jenkins et al., 2018; Nguyen, 2017). It is possible that some type of whole-child curriculum may be particularly effective for boosting children's school readiness, there is currently little to no evidence to support that conclusion.

The data used in this study represent the business-as-usual classroom experiences of low-income preschoolers and these results do not represent tests of efficacy of various curricula because they may have been questionably implemented in the classroom. However, the classrooms included in the FACES study may likely reflect the degree of implementation that is actually found in classrooms. One limitation to this study is that the data do not contain measures for whether curricula were implemented with high fidelity in the classroom or whether and what kind of extensive professional development teachers received. Teachers may report using a particular curriculum that they may reference on occasion, or not at all, for their daily lesson plans. Though multiple controls and grantee fixed effects were included, these results are descriptive in nature and cannot be interpreted as causal, nor can they be generalized to the national population of children.

Curricula, including those most widely used in publicly funded preschool centers, may be much more beneficial for children's school readiness if teachers are provided adequate supports

and development opportunities to fully implement the curricular activities. A post-hoc review of the teacher survey items about classroom curricula indicated that teachers receive some initial training on the curricula and that they like the curriculum they use, but they do not receive consistent and continuous support in implementing the curriculum. Fidelity of curricular implementation may be a vital component of influencing children's outcomes and should be a direction of future research. Given that most publicly funded programs are using one of the whole-child curricula included in the current study, and it is not yet known how consistently curricula are implemented across and within these centers, one important future research direction is to further understand the link between implementation and effectiveness. Central to this work is what can be done to support teachers and ensure consistency across Head Start centers and classrooms, especially when public programs are scaled up (Sarama, Clements, Wolfe, & Spitler, 2012).

The current curricular investments in early childhood policy need to be carefully scrutinized and reconsidered. It is important to critically evaluate the relative effects of curricular packages on key classroom dimensions and school readiness outcomes, particularly for children who are disadvantaged and to whom many programs are targeted. Considering the average price tag of a curriculum package per classroom—upwards of $2,000—there is a need for more research on whole-child curricula more generally, and on Creative Curriculum and High Scope specifically given their prevalence in Head Start. Given the large investments and the high-stakes accountability pressures Head Start faces, getting the developmental and academic content right in public early learning programs is absolutely critical to promoting school readiness for our most at-risk preschoolers.

References

Achenbach, T. M., Edelbrock, C., & Howell, C. T. (1987). Empirically based assessment of the behavioral/emotional problems of 2- and 3- year-old children. *Journal of Abnormal Child Psychology, 15*(4), 629-650. doi:10.1007/bf00917246

Alexander, K. L., Entwisle, D. R., Blyth, D. A., & McAdoo, H. P. (1988). Achievement in the first 2 years of school: Patterns and Processes. *Monographs of the Society for Research in Child Development, 53*(2), i-157. doi:10.2307/1166081

Belfield, C. R., Nores, M., Barnett, W. S., & Schweinhart, L. J. (2006). The High/Scope Perry preschool program. *Journal of Human Resources, XLI*(1), 162-190.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, *57*(2), 111-127.

Clements, D. H. (2007). Curriculum research: toward a framework for "research-based curricula." *Journal for Research in Mathematics Education*, 35-70.

Clifford, R. M., Barbarin, O., Chang, F., Early, D. M., Bryant, D., Howes, C., et al. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science, 9*(3), 126-143.

Copple, C., & Bredekamp, S. (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8* (3rd ed.). Washington, DC: National Association for the Education of Young Children.

DeVries, R., & Kohlberg, L. (1987). Programs of early education: the constructivist view. White Plains, NY: Longman.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do." *Developmental Psychobiology*, *29*(4), 315-334.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third Edition (PPVT-III)*. Upper Saddle River, NJ: Pearson Publishing.

Fuligni, A. S., Howes, C., Huang, Y., Hong, S. S., & Lara-Cinisomo, S. (2012). Activity settings and daily routines in preschool classrooms: Diverse experiences in early learning settings for low-income children. *Early Childhood Research Quarterly, 27*(2), 198-209.

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System (SSRS)*. Circle Pines, MN: American Guidance Service.

Hulsey, L. K., Aikens, N., Kopack, A., West, J., Moiduddin, E., & Tarullo, L. B. (2011). *Head Start children, families, and programs: Present and past data from FACES.* Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Jenkins, J. M., & Duncan, G. J. (2017). Do pre-kindergarten curricula matter? In D. Phillips, K. A. Dodge, & Pre-Kindergarten Task Force (Eds.), *The current state of scientific knowledge on pre-kindergarten effects* (pp. 37-44). Washington, DC: Brookings Institution and Duke University.

Jenkins, J.M., Duncan, G.J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. (2018). Boosting school readiness: Should preschools target skills or the whole child? *Economics of Education Review*. doi: 10.1016/j.econedurev.2018.05.001

National Center on Quality Teaching and Learning. (2014). *Preschool Curriculum Consumer Report.* Washington, DC: Office of Head Start, Administration for Children and Families, U.S. Department of Health and Human Services.

Nguyen, T. (2017). Impacts of targeted and global preschool curricula on children's school readiness: A meta-analytic review. Working Paper, Irvine Network on Interventions in Development.

Phillips, D. A., Gormley, W. T., & Lowenstein, A. E. (2009). Inside the pre-kindergarten door: Classroom climate and instructional time allocation in Tulsa's pre-K programs. *Early Childhood Research Quarterly, 24*(3), 213-228. doi.org/10.1016/j.ecresq.2009.05.002

Piaget, J. (1976). *Piaget's theory.* New York, NY: Springer.

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

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness*. (NCER 2008-2009). Washington, DC: National Center for Education Research, Institute for Education Sciences, U.S. Department of Education.

Sarama, J., Clements, D. H., Wolfe, C. B., & Spitler, M. E. (2012). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies. *Journal of Research on Educational Effectiveness, 5*(2), 105-135.

Schweinhart, L. J. (2005). *Lifetime Effects: The High/Scope Perry preschool study through age 40.* Ypsilanti, MI: High/Scope Educational Research Foundation.

U.S. Department of Education. (2013). *Early childhood education intervention report: The Creative Curriculum for Preschool, Fourth Edition*. Washington, DC: Institute of Education Sciences, What Works Clearinghouse.

U.S. Department of Health and Human Services. Administration for Children and Families. Office of Planning, Research and Evaluation. Head Start Family and Child Experiences Survey (FACES): 2009 Cohort. (2009-2013). ICPSR34558-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. doi:10.3886/ICPSR34558.v1

Weikart, D. P., & Schweinhart, L. (1987). The High/Scope cognitively oriented curriculum in early education. In J. L. Roopnarine & J. E. Johnson (Eds.), *Approaches to early childhood education* (pp. 253-268). New York: Merill/Macmillan.

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

Zill, N. (1990). *Behavior problems index based on parent report*. Washington, DC: Child Trends.

Table 2.1. Descriptive statistics of child and family characteristics, classroom quality measures, teacher characteristics, and classroom and center characteristics by curriculum.

	High Scope	Creative Curriculum	High Reach	Scholastic	Other Published Curricula	Diff
Child and Family Characteristics						
White	0.16	0.26	0.02	0.25	0.16	*
African-American	0.42	0.26	0.16	0.40	0.35	*
Hispanic	0.31	0.40	0.77	0.13	0.41	*
Asian or Pacific Islander	0.03	0.02	0.01	0.00	0.02	
Other race/ethnicity	0.08	0.06	0.03	0.21	0.06	*
Female	0.43	0.52	0.51	0.59	0.49	
Standardized baseline assessments						
PPVT	0.05	0.01	-0.40	0.28	-0.05	*
WJ Letter Word	0.07	-0.03	-0.10	0.22	0.00	
WJ Spelling	0.05	-0.03	0.12	0.19	-0.03	
WJ Applied Problems	-0.01	0.02	0.00	0.03	-0.02	*
Behavior Problems	0.09	-0.03	0.03	-0.34	0.05	
Social Skills	-0.11	0.02	0.12	0.10	-0.08	
Maternal Education						
Less than high school diploma	0.31	0.33	0.36	0.18	0.34	
High school diploma or GED	0.30	0.31	0.23	0.27	0.30	
Some college degree	0.22	0.21	0.23	0.30	0.18	*
Bachelor's degree or higher	0.07	0.04	0.06	0.12	0.07	*
Observations (Children)	424	1699	179	106	709	
Classroom Quality						
Total ECERS Score	4.14	4.45	4.16	3.60	4.17	*
Factor 1 Language/Interactions	4.59	4.81	4.72	3.91	4.63	*
Factor 2 Provisions for Learning	3.79	4.18	3.65	3.31	3.76	*
Total CLASS Score	3.62	3.80	3.81	3.60	3.73	*
Concept Development	1.98	2.10	1.93	2.08	2.15	*
Quality of Feedback	2.07	2.34	2.13	2.21	2.29	*
Language Modeling	2.25	2.52	2.68	2.05	2.45	*
Positive Climate	5.14	5.40	5.38	5.64	5.24	*
Negative Climate	1.40	1.18	1.15	1.59	1.22	*
Teacher Sensitivity	4.53	4.75	4.62	4.40	4.64	*
Regard of Student Perspectives	4.39	4.54	4.56	4.06	4.46	*
Behavior Management	4.93	5.13	5.18	4.71	4.96	*
Productivity	4.75	4.99	5.06	4.67	4.86	*
Instructional Learning Formats	3.83	4.07	4.41	3.78	4.05	*
Instructional Support	2.10	2.32	2.24	2.11	2.30	*
Emotional Support	5.17	5.38	5.35	5.13	5.28	*
Classroom Organization	4.50	4.73	4.89	4.38	4.62	*
Teacher Characteristics						
Female	0.97	0.99	1.00	1.00	1.00	
Annual Salary	26841.69	26604.35	21956.52	17636.36	30263.62	*
Years teaching Head Start	9.33	8.91	9.77	9.49	8.68	
Years of teaching experience	13.75	13.95	13.22	11.73	12.64	
Child Development Associate's	0.50	0.50	0.65	0.59	0.41	*
Teaching Certificate or License	0.35	0.44	0.23	0.00	0.53	*
High school of less	0.06	0.05	0.17	0.15	0.06	^
Some college	0.07	0.12	0.21	0.17	0.08	^
Associate's	0.40	0.37	0.51	0.24	0.29	^
Bachelor's	0.40	0.35	0.06	0.36	0.39	^
More than a Bachelor's	0.07	0.12	0.05	0.07	0.18	^

Classroom and Center Characteristics						
Class size in observed class Spring	17.03	17.00	17.55	16.57	17.45	^
3 years old or younger	0.49	0.44	0.39	0.49	0.43	*
4 year olds	0.47	0.49	0.57	0.48	0.51	*
5 year olds	0.04	0.07	0.05	0.03	0.06	*
Length of school day in hours	5.91	5.21	5.08	5.78	5.94	*
Observations (Classrooms)	66 (14%)	251 (55%)	24 (5%)	14 (3%)	101 (22%)	

Note. * p<.05 from ANOVA. ^p<.05 from Pearson Chi-Squared test. Baseline assessments are standardized to mean=0, standard deviation=1. Other curricula category included: Let's Begin with the Letter People, Montessori, Bank Street, Creating Child Centered Classrooms - Step by Step, Curiosity Corner - Johns Hopkins, and locally designed curricula. Weight used = CNST1WT for child and family characteristics; T1TCHWT for classroom quality and teacher characteristics; C1WT for center characteristics.

Table 2.2. Regression results for classroom quality outcomes for classrooms using other published curricula compared with Creative Curriculum.

	ECERS		CLASS	
	(1)	(2)	(1)	(2)
Other Published Curricula	-0.23+	-0.22	0.16	0.22
	(0.13)	(0.25)	(0.13)	(0.26)
Grantee fixed effects		X		X
Observations	451	451	457	457

Note. Creative Curriculum is the reference group. Standard errors in parentheses and clustered at the program level. Statistics were weighted to represent all children entering Head Start for the first time in the cohort year. Other curricula category included: Let's Begin with the Letter People, Montessori, Bank Street, Creating Child Centered Classrooms - Step by Step, Curiosity Corner - Johns Hopkins, and locally designed curricula. Model covariates included: teachers' education, race, years of experience, and classroom-level aggregates of children's race, gender, and parental education. Weight used = T1TCHWT. +p<.10; *p<.05; **p<.01.

Table 2.3. Regression results for associations between various curricula and classroom quality.

	ECERS		CLASS	
	(1)	(2)	(3)	(4)
High Scope	-0.19	-0.14	-0.13	0.38
	(0.24)	(0.28)	(0.23)	(0.38)
High Reach	-0.34*	-0.31*	-0.29	0.13
	(0.16)	(-0.18)	(0.34)	(0.43)
Scholastic	-0.34+	-0.30	0.23	0.12
	(0.18)	(0.20)	(0.15)	(0.30)
Other Published Curricula	-0.32+	-0.28	-0.04	0.17
	(0.19)	(0.21)	(0.14)	(0.31)
Grantee fixed effects		X		X
P-value for joint significance	0.06	0.04	0.34	0.95
Observations	451	451	457	457

Note. Creative Curriculum is the reference group. Standard errors in parentheses and clustered at the program level. Statistics were weighted to represent all children entering Head Start for the first time in the cohort year. Other curricula category included: Let's Begin with the Letter People, Montessori, Bank Street, Creating Child Centered Classrooms - Step by Step, Curiosity Corner - Johns Hopkins, and locally designed curricula. Model covariates included: teachers' education, race, years of experience, and classroom-level aggregates of children's race, gender, and parental education. Weight used = T1TCHWT. +p<.10; *p<.05; **p<.01.

Table 2.4. Regression results for child outcomes for classrooms using other published curricula compared with Creative Curriculum.

	PPVT		WJ Applied Problems		WJ Letter Word		WJ Spelling		Behavior Problems		Social Skills		Pencil Tapping Task	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Other Published Curricula	-0.02	-0.01	-0.02	-0.05	-0.04	-0.01	0.06	0.08	-0.17	-0.14	-0.10	0.00	0.09	0.03
	(0.05)	(0.10)	(0.08)	(0.10)	(0.08)	(0.11)	(0.08)	(0.10)	(0.15)	(0.16)	(0.18)	(0.20)	(0.09)	(0.13)
Grantee fixed effects		X		X		X		X		X		X		X
Observations	2611	2611	2397	2397	2401	2401	2477	2477	2691	2691	2736	2736	1473	1473

Note. Creative Curriculum is the reference group. Standard errors in parentheses and clustered at the program level. Statistics were weighted to represent all children entering Head Start for the first time in the cohort year. Other curricula category included: Let's Begin with the Letter People, Montessori, Bank Street, Creating Child Centered Classrooms - Step by Step, Curiosity Corner - Johns Hopkins, and locally designed curricula. Pencil tapping task was only administered to children 4 years or older. All models are residualized change. Child and family characteristic covariates include: child race, child gender, child age, mother's education, family income, whether mother is working, and mother's age. All outcomes are standardized. PPVT=Peabody Picture Vocabulary Test, WJ=Woodcock-Johnson. Weight used = CNST1WT. Coefficients for these models each come from separate regression models. +p<.10; *p<.05; **p<.01.

Table 2.5. Regression results for associations between various curricula and child outcomes.

	PPVT		WJ Applied Problems		WJ Letter Word		WJ Spelling		Behavior Problems		Social Skills		Pencil Tapping Task	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
High Scope	-0.01	-0.05	-0.03	-0.04	0.03	0.18	-0.01	-0.01	-0.22	-0.27	-0.18	-0.06	-0.09	-0.10
	(0.04)	(0.05)	(0.08)	(0.09)	(0.09)	(0.12)	(0.04)	(0.06)	(0.19)	(0.25)	(0.22)	(0.26)	(0.16)	(0.19)
High Reach	-0.30*	-0.33**	-0.25**	-0.18*	-0.17+	-0.11	-0.11	-0.24	0.03	0.09	-0.11	-0.29+	-0.04	-0.02
	(0.12)	(0.10)	(0.05)	(0.08)	(0.09)	(0.13)	(0.13)	(0.16)	(0.09)	(0.13)	(0.16)	(0.17)	(0.10)	(0.14)
Scholastic	-0.12	-0.13	0.03	0.04	-0.06	-0.04	-0.07	-0.02	-0.23	-0.29	-0.02	-0.09	-0.16	-0.20
	(0.14)	(0.16)	(0.07)	(0.08)	(0.06)	(0.08)	(0.07)	(0.15)	(0.14)	(0.18)	(0.12)	(0.23)	(0.26)	(0.27)
Other Published Curricula	0.05	0.04	0.06	0.12+	0.09	0.04	0.07	0.06	-0.06	-0.15	0.10	0.16	0.09	-0.01
	(0.03)	(0.06)	(0.06)	(0.07)	(0.07)	(0.09)	(0.07)	(0.07)	(0.09)	(0.10)	(0.09)	(0.11)	(0.05)	(0.08)
Grantee fixed effects		X		X		X		X		X		X		X
P-value for joint significance	0.00	0.02	0.04	0.05	0.09	0.18	0.41	0.62	0.73	0.81	0.33	0.25	0.51	0.76
Observations	2611	2611	2397	2397	2401	2401	2477	2477	2691	2691	2736	2736	1463	1463

Note. Creative Curriculum is the reference group. Standard errors in parentheses and clustered at the program level. Statistics were weighted to represent all children entering Head Start for the first time in the cohort year. Other curricula category included: Let's Begin with the Letter People, Montessori, Bank Street, Creating Child Centered Classrooms - Step by Step, Curiosity Corner - Johns Hopkins, and locally designed curricula. Pencil tapping task was only administered to children 4 years or older. All models are residualized change. Child and family characteristic covariates include: child race, child gender, child age, mother's education, family income, whether mother is working, and mother's age. All outcomes are standardized. PPVT=Peabody Picture Vocabulary Test, WJ=Woodcock-Johnson. Weight used = CNST1WT. Coefficients for these models come from five separate regression models. +p<.10; *p<.05; **p<.01.

CHAPTER THREE

Study 2: High Scope Curriculum Implementation in Head Start Classrooms: A Qualitative Evaluation

Commonly used early childhood curricula, which most typically follow a model of educating the whole child, have come under great scrutiny in recent years (e.g. Jenkins & Duncan, 2017). Experimental (Jenkins et al., 2018) and meta-analytic (Nguyen, 2017) results from recent studies point to the relative ineffectiveness of whole-child curricula. Despite their widespread adoption, these studies suggest that whole-child curricula are no more effective at boosting children's school readiness than the assortment of activities that early childhood education centers develop on their own. The evidence on these curricula as it stands is concerning for Head Start programs, given that they are all mandated to use a whole-child curriculum in their classrooms.

The purpose of the current study is to follow up on the quantitative findings from previous studies to shed light on how this type of early childhood curriculum is implemented in practice across different Head Start classrooms. Below, I discuss what is known about whole-child curricula, evidence of their effectiveness (or lack thereof) from the literature, and why a qualitative study of fidelity implementation is needed, followed by a description of the study framework and a whole-child curriculum and its components that is the focus of this study.

Whole-Child Curricula

Performance standards for Head Start require that all grantees implement curricula educate the "whole-child," are developmentally appropriate, and encourage classroom interactions that promote children's development in multiple domains (U.S. Department of Health & Human Services, 2016; Zigler & Bishop-Josef, 2006). Survey data from the 2009 Head

Start Family and Child Experiences Survey (FACES) show that the majority of Head Start teachers report using Creative Curriculum for Preschool (53%) and High Scope Curriculum (15%) as their primary curriculum (Moiduddin, Aikens, Tarullo, West, & Xue, 2012). Apart from these two widely used curricula packages, teachers have also reported using other published whole-child curricula (e.g., Scholastic, High Reach). Whole-child or "global" curricula are not specific to an academic content domain but rather promote children's learning through a developmental-interaction approach. These curricula are based on Piagetian theory, which emphasizes child-centered active learning that is cultivated through strategic arrangement of the classroom environment (DeVries & Kohlberg, 1987; Piaget, 1987; Weikart & Schweinhart, 1987), and sociocultural theory, where the teacher provides supportive and responsive interactions with children (Vygotsky, 1978). Children are encouraged to interact independently with the equipment, materials, and other children in the classroom environment, and teachers support children's critical thinking and problem-solving by provided open-ended learning opportunities to practice these skills.

Though whole-child curricula have intuitive appeal, results from a recent meta-analysis by Nguyen (2017) indicate that they are no more effective in boosting children's school readiness outcomes than locally-developed curricula, which include materials that teachers and programs develop on their own. Specifically, the meta-analysis reports an insignificant and small effect size of .08 SD on children's academic outcomes (measured as a literacy, language, and math composite). This evidence is corroborated with work by Jenkins et al., (2018) who also found statistically insignificant differences between the school readiness skills of children exposed to a whole-child curriculum (Creative Curriculum) and locally-developed curricula, with a small point estimate of .02 SD. It is possible that the variability in which the curriculum is

implemented contributes to these null findings. However, neither the Jenkins et al. (2018) nor the Nguyen (2017) studies are ideal for studying whole-child curriculum implementation because either the curricula were not implemented with high fidelity per the developer's recommended conditions (Jenkins et al., 2018) nor were fidelity of implementation measures were not coded (Nguyen, 2017). Making sense of these null findings requires an understanding of how these curricula are implemented in real life at scale, which is the goal of the current study.

Approaches to Studying Curriculum Implementation

Researchers have increasingly paid attention to fidelity of implementation—or the degree to which teachers implement a program as intended by the developers—to better understand the impact of research-based curricula on child and classroom outcomes (Odom, 2009; Odom et al., 2010), but such studies in early childhood are still at a relatively early stage (Mendive, Weiland, Yoshikawa, & Snow, 2016). In an attempt to capture implementation in large-scale datasets such as FACES, teacher survey items ask about their attitudes towards the curriculum, whether they have training in the curriculum, whether they have the necessary materials to implement the curriculum, and whether the curriculum leaves room for teacher creativity. However, these survey items ask little about the quantity and quality of curriculum content. In the early childhood curriculum literature, researchers have applied O'Donnell's (2008) classification of implementation as 'structural' or 'process' to their measurement of implementation, similar to Dane and Schneider's (1998) dimensions of *dosage* (how much of the program was actually delivered) and *adherence* (level of correspondence between the intended intervention and the version implemented).

Structural measures of implementation rely on assessing implementation by the amount of instruction delivered or number of lessons provided. For example, in the Infant Health and

Development Project (IHDP) designed to promote the development of children born with low birth weights implemented *Early Partners* for its youngest participants and *Partners for Learning* curricula (Ramey et al., 1992). An index of participation, including the number of home visits, parent meetings attended, and child attendance in centers, was used to assess implementation. Further, teachers' implementation of curricula in the *Reading First* initiative was assessed by the amount of reading instruction that was delivered during the day (Al Otaiba et al., 2008). Overall, these structural measures of implementation aim to capture the delivery of the quantity of the specified curriculum content.

In the case of process measures of implementation, researchers typically use observations of teachers in classrooms and checklists or ratings that capture teachers' delivery of the key instructional components of the curriculum. Pence, Justice, and Wiggins (2008) evaluated the implementation of the *Language Focused Curriculum* by using a procedural checklist during their classroom observations measured three times over the academic year. This checklist included 45 items that trained observers would conduct over a two-hour class period, rating the core curriculum components either as a dichotomous rating (present or not) or on a scale of 0-3 (where each point represented the number of times a component was observed). Similarly, Bierman et al., (2008) assessed implementation of the Head Start REDI curriculum by having trainers come in for monthly observations to rate the quality of implementation of the four program components, in addition to an overall rating of REDI implementation, using a 6-point Likert scale going from 'poor' to 'exemplary.'

Though these process measures for fidelity of implementation provide a richer level of detail about the delivery of the curriculum content than the survey items and structural implementation measures, the data collected with these measures are reduced to a crude variable

that may not allow for flexibility of curriculum implementation or adequately capture the complexities and idiosyncratic aspects of classroom processes and interactions that contribute to fidelity. Because no set of variables is complete or appropriate for each situation, qualitative approaches can supplement these analyses to meaningfully document implementation processes and the key variables that affect it (Clements, 2007; Greene, 2015; Lincoln & Guba, 1985). The current study is the first to my knowledge to use open-ended qualitative observations of implementation fidelity of a whole-child curriculum in early childhood classrooms.

Theory-Driven Evaluation Approach

This study is grounded in a theory-driven evaluation framework (Chen, 1990; 2005; Chen & Rossi, 1983) in which the theory behind a program is mapped out and then a research evaluation is designed to test that theory as a means to understand when and how the program works in different settings (Weiss, 1997). The program theory is the implicit set of assumptions about how an intervention is meant to work and what impacts it is expected to have. Based on the program theory, researchers can design an evaluation that systematically examines how these assumptions operate in the real world. This framework allows for researchers to explore many more factors outside the typical evaluation research focus (Chen, 2005), such as contextual factors, adaptations, implementation facilitators and barriers, and thresholds of fidelity. Results are then used for program development and improvement rather than as evidence for the effectiveness of the program itself (Patton, 2014). Specifically, the aim of theory-driven evaluation work is to enable decision-makers to reach a deeper understanding of the program and how it can be made to work most effectively. In the current study, I outline and describe the program theory of the High Scope curriculum and then qualitatively evaluate how teachers tailor

and implement it with fidelity and/or ways in which implementation was limited or absent when applied to different children and classroom settings.

The High Scope Curriculum

The most famous example of a preschool program based on a whole-child curriculum is the Perry Preschool study, which used the High Scope curriculum (Belfield, Nores, Barnett, & Schweinhart, 2006; Schweinhart, 2005). The only evidence of High Scope comes from an experiment conducted in the 1960s with 123 children, which found that it improves children's early-grade cognitive scores and reduces early adult outcomes such as crime. However, the results from this experiment are difficult to generalize to today's early childhood education landscape because the counterfactual conditions likely do not apply today (i.e., children who did not attend center-based preschool and very low levels of maternal schooling).

The curriculum today has gone through a few iterations since Perry, but still remains true to the whole-child philosophy (Epstein & Hohmann, 2012; High Scope Educational Research Foundation, personal communication, August 7, 2017). French (2012) notes that "High Scope is continually being updated to reflect the latest child development research" (p. 133). High Scope is based on the philosophy of *active participatory learning*, in which "teachers and students are partners in shaping the learning experience" (Epstein, 2014, p. 8). Learning occurs in a play-based environment where children and adults work together to develop children's knowledge and skills in problem solving and decision-making, and to promote their sense of curiosity, persistence, and creativity. A large part of this active learning process is influenced by how the classroom is set up, what materials are made available for children, the daily routine, and the curriculum content—referred to as "Key Developmental Indicators" (KDIs). These KDIs provide the content that is incorporated throughout each of the High Scope curriculum components.

In High Scope classrooms, the learning environment is purposefully divided into well-defined interest areas with places for group and individual activities. The classroom is labeled so that children can find, use, and return materials and relate the printed word to materials that interest them. The classroom is filled with materials that support a wide range of play experiences and reflect the children's family lives. There are also many real and natural materials as well as open-ended materials so that children can develop their imagination.

The curriculum consists of four main components: (1) engaging children in planning, enacting, and reviewing their free play time (i.e., "plan-do-review" or "planning time"); (2) focusing their lesson plans and activities on children's interests; (3) incorporating all school readiness learning domains; and (4) scaffolding children's learning. These components are integrated into the High Scope daily routine of large-group time, small-group time, planning time, work time (i.e., free play), cleanup time, recall time, outside time, and snack or lunch time every day. The High Scope curriculum places a strong emphasis on the need for a consistent daily routine for children. Components of the curriculum are not mutually exclusive and multiple elements can be observed happening at the same time in the classroom. These components are described in further detail below.

Plan-do-review. "Plan-do-review" is an important component of the High Scope routine as children do this every day. High Scope calls these parts of the day *planning time, work time,* and *review (or recall) time* (Epstein, 2014). This process provides children an opportunity to think through and make choices about what they are going to do, enact those ideas, and then reflect on their activities. During planning time, children describe their plans to the teacher and other children in their group. The teacher responds to the child's ideas and might suggest ways to enhance their plans. For example, if a child plans to throw a party in the house area, the teacher

may ask what type of the party it will be and what materials they will need. After discussing their plans, children are given ample time to "do" or "work" on them during work time (e.g., free play time). This work time is always of children's choosing as it is an opportunity for them to freely play and explore around the classroom. High Scope does not recommend that teachers task children with doing something or send them to a particular area of the classroom during this time. After working, all children actively participate in cleaning up. To engage children in reviewing what they did during their free play time, the teacher will have the children reconvene back to their groups for recall time. During recall time, children explain to the teacher what they did, how they did it, and with whom they worked, and discuss any problems they might have had. Plan-do-review is explicitly built into the daily High Scope routine. Through this process, High Scope teachers help children "gain confidence as thinkers, problem solvers, and decision makers" (Epstein, 2014, p. 19).

Child-centered active learning. High Scope considers active learning to be the cornerstone of their approach to educating young children, in which they are provided with direct, hands-on experiences with people, objects, events, and ideas (Epstein, 2014; Epstein, Schweinhart, & McAdoo, 1996; Wiltshire, 2012). Active learning is largely based on children's interests and choices. Abundant materials are provided that appeal to children's senses and can be used in a variety of ways. High Scope argues that learning occurs through children's direct actions on and thinking about the materials. Their cognitive and socio-emotional development often occurs through interactions with physical objects. Since learning results from children's attempts to pursue their own interests and goals, the opportunity for children to choose activities and materials is considered essential (Epstein, 2014). Open-ended materials are preferred in High Scope as they can help promote children's language growth, imagination, and problem-solving

skills each time children interact with them (Epstein, 2014), Through their language and thought, children are encouraged to reflect and communicate what they are doing. They may also seek cooperation of their peers and adults in the classroom in their activities. With some guidance and scaffolding from the teacher, children develop their reasoning, problem-solving, and creativity.

Learning across all domains. Because High Scope is a comprehensive whole-child curriculum, it focuses on promoting children's learning across all developmental domains through the use of KDIs. These domains include approaches to learning; social and emotional development; physical development and health; language, literacy, and communication; mathematics; creative arts; science and technology; and social studies (Epstein, 2014; 2016). Each of these domains includes KDIs that define the knowledge children should be able to acquire in an early learning setting (Epstein, 2014, p.12). For example, language, literacy, and communication has nine KDIs that include comprehension, speaking, vocabulary, phonological awareness, alphabetic knowledge, reading, concepts about print, book knowledge, and writing. Figure 3.1 provides the entire list of High Scope KDIs. High Scope leaves the specific details of the lesson planning open-ended for teachers and instead encourages them to use the KDIs as a guideline for planning the different classroom activities based on children's interests and abilities (Epstein, 2014; High Scope Educational Research Foundation, personal communication, August 10, 2017).

[Insert Figure 3.1]

Scaffolding children's learning. The High Scope curriculum encourages teachers to scaffold learning by supporting the child's current level of understanding and gently extending their thinking and reasoning to the next level (Epstein, 2014). This type of "light-touch instruction" occurs throughout the day in all interactions and activities. Scaffolding consists of an

adaptive strategy that recognizes the current capacities of individual children and guides him or her to extend their learning without too much frustration. Interactions and activities are individually tailored to the child's current ability level and it must be neither too hard nor too easy in order to keep them in their "zone of proximal development," or the level of difficulty at which the child can learn the most. For example, during small-group time children might practice recognizing the letters of their name on their badge. Once a child is able to recognize specific letters, the teacher can teach them the sound and then eventually work together to think of different words that start with that sound. High Scope encourages teachers to first validate, or support, what children already know, and then when the time is right, gently encourage them to extend their thinking to the next level.

Current Study

The aim of the current study is to get inside the "black box" of whole-child curricula and qualitatively evaluate fidelity of curriculum implementation across four Head Start classrooms. This study was guided by a theory-driven evaluation approach as described by Chen (2005). Given the flexible nature of data collection procedures utilized in theory-driven evaluation (Chen, 1997; Chen & Rossi, 1992), qualitative data collection methods were used to answer the key research question. Because the High Scope curriculum is driven largely by child-centered and teachers adapt their lesson plans accordingly, I focus on the process implementation of curriculum fidelity that O'Donnell (2008) describes. The focal phenomenon of these observations was the way teachers implemented the four main components of the High Scope curriculum by engaging children in planning, enacting, and reviewing their free play time (i.e., "Plan-Do-Review"), focusing their lesson plans and activities on children's interests, incorporating all school readiness learning domains through the use of KDIs, and scaffolding

children's learning. The key research question driving this study is: how do teachers implement a whole-child curriculum across different classrooms?

Method

Context

Head Start is a federally funded preschool program dedicated to improving the health, well-being, and school readiness for 3- to 5-year old children and their families living in poverty. The Federal Office of Head Start contracts with local grantee agencies throughout the country to implement Head Start by providing center-based and home-based early childhood education in addition to health, nutrition, and other comprehensive services. The local Head Start umbrella agency that is the focus of this study is located in a large, urban county in a Southwestern region of the U.S. It is currently one of the largest Head Start grantee agencies in the region, providing services to nearly 4,000 infants, toddlers, preschoolers, pregnant mothers, and families, including homeless and foster families and teen parents. Among the children in this agency, 81% are Hispanic, 8% Asian, 7% White, and 1% Black. Approximately 50% of children are dual language learners. Sixty-three percent of teachers are Hispanic, 15% White, 12% Asian, and 4% Black.

All classrooms are required to use the High Scope curriculum, and this policy has been in place for more than a decade (Educational Director, personal communication, December 13, 2016). The agency does not have clear guidelines for how teachers are trained or provided with professional development to implement the High Scope curriculum specifically. However, the Educational Director noted that their teachers are very familiar with the curriculum and its philosophy and that the agency conducts ongoing assessments of teaching practices throughout the year (Educational Director, personal communication, December 13, 2016). This agency has a

history of working together with researchers from the local university and the directors were actively engaged in research-practice partnerships at the time of this study (Executive Director, personal communication, May 23, 2016). In particular, I have an existing multiyear relationship with this agency through research visits and community events, which has helped me to develop familiarity with the directors, teachers, and other staff members.

Participants

In order to sample teachers, I met with the Educational Director to discuss the minimum criterion for participation, which was some degree of familiarity and experience with implementing the High Scope curriculum in their classrooms. The Educational Director made suggestions for which centers to observe, and then I went to each center where the center's Director placed me into a classroom. Four classrooms, all located in different Head Start centers across the region, were chosen as a number that would sufficiently allow for the detailed level of observational field notes.

Each classroom had a lead teacher and assistant teacher. Thus, I observed a total of eight teachers across the four classrooms. The teachers ranged in age from their mid-30s to late-40s and were all Hispanic females. Teachers had 12-18 years of education ($M = 13.77$; $SD = 1.48$) with two lead teachers having a Bachelor's degree or higher, and 8-25 years of experience teaching preschool ($M = 16.52$; $SD = 8.49$). All of the lead and assistant teachers had been working with the agency for seven years or more, except for one assistant teacher who recently joined the agency less than a year ago at the time I conducted the observations. Classrooms included 12-17 children, 48% boys, with an average age of 51.33 months ($SD = 5.75$).

I received Institutional Review Board approval for this study. Teachers provided written consent after receiving verbal and written explanations of the following: the overall aim of the

study, the voluntary nature of participation, their right to withdraw consent to participate at any time, including after the study's conclusion, assurance of anonymity, and the use of data for no other purpose. Observations were carried out only after the teachers consented. Pseudonyms are used throughout the study for all of the participants in an effort to protect their confidentiality.

Data Collection

I conducted open-ended observations of curriculum implementation fidelity across four Head Start classrooms. Since teachers planned their lessons out weekly, I spent a week in each classroom (five days) to observe how their plans unfolded. All classrooms I observed were half-day programs. I spent the entire time (8:30 AM to 11:30 AM or 1:00 PM to 4:00 PM) each day in the classrooms to ensure that I had the opportunity to see all of the recurring features of each teacher's typical working day. The daily schedule can be found in Appendix Table 3.1. As a participant observer, I was active in the classroom, assisting in activities when needed or supporting various classroom events. When teachers interacted with children, I watched and recorded these observations but did not ask them about it in the moment. At the end of each day of observation, I had informal conversations with the teachers about the curriculum and the activities I observed. Their responses to these questions focused my future observations, which in turn led to more questions and more informal conversations.

Daily field notes were recorded by hand on the spot and expanded shortly after the observation. These observational notes included documentation of the events and interactions of the day, such as small- and large-group activities, individual activities, literacy time, free play, and adult-child interactions. Approximately 15 hours of observational data were collected from each teacher's classroom, amounting to a total of 60 observation hours. Collection of copies of teachers' lesson plans and other classroom artifacts and meetings with the center directors

provided additional detail and triangulation of the data. The findings and conclusions for this study are based primarily data from field notes, transcripts of conversations, and copies of lessons plans. Additionally, I made copies of the teachers' guide used during the more formal teacher-directed activities during small-group time. I exchanged emails with the educational director, center directors, and teachers to follow up with them as needed, and these emails were used as member checks to enrich the interpretations of the observational data. Prolonged engagement was achieved through the in-depth observations and through the extensive time that I spent working with the Head Start agency that supported this study.

Following the observations, I received 30 hours of in-person training on the High Scope curriculum by the developers themselves at the High Scope Educational Research Foundation. I was provided with an overview of the components of the curriculum, including the curriculum content areas, the High Scope daily routine, effective adult-child interaction strategies, active participatory learning, scaffolding children's learning, and lesson planning. This training provided me with a better understanding and context for what I was observing in the classrooms and ways in which my own observations converged or diverged from what High Scope considered ideal implementation.

Data Analysis

As an assessment of fidelity of curriculum implementation, I identified *a priori* the components of High Scope from the curriculum manual (Epstein, 2014) to deductively code my data. Deductive analysis was the most appropriate method given that study goal was to test whether the observational classroom data gathered were consistent with the curriculum components. This process allows for researchers to examine their data with a more narrowed focus, using predetermined coding categories based on the research questions (Elo & Kyngäs,

2008; Hsieh & Shannon, 2005). The observational field notes were sorted into different categories based on the four main components of the High Scope curriculum—plan, do, review; learning across all domains; child-centered active learning; and scaffolding—and was used as a lens in the analysis of the observations. These components have explicit definitions, and the coding frame was designed on the basis of the definitions of each component in Table 3.1. The different High Scope components interact, and some items were related to more than one component. In those cases, the field notes were coded with the component that was deemed to be the most relevant. I shared the coding with the Head Start teachers and center directors, and also consulted with the High Scope Early Child Specialist and Demonstration Preschool Teacher to ensure that the field notes were accurately categorized based on the descriptions of the components. Finally, I wrote a range of memos throughout data collection and analysis for each classroom in order to keep a clear and detailed account of the decisions made during the data analysis (Creswell & Miller, 2000).

Findings

Below, I return to the four main components of the High Scope curriculum and describe the observational findings from most to least occurring in the classrooms: (1) engaging children in plan-do-review; (2) focusing lesson plans and activities on children's interests; (3) incorporating all school readiness learning domains; and (4) scaffolding children's learning. Classroom examples of each curriculum component are provided in Table 3.1. The examples presented in the table were developed with assistance from the High Scope Early Child Specialist and Demonstration Preschool Teacher, so portions of the examples were actually observed in classrooms while some were not.

[Insert Table 3.1]

Plan-Do-Review

The plan-do-review process is the heart of the High Scope daily routine and is explicitly outlined in all of the teachers' lesson plans. Teachers and children all knew that plan-do-review was something they did every day when children communicate their ideas of where they will work, who they will work with, what materials they want to use, what they want to do, and how they will go about doing it. I observed the four teachers and their assistant teachers implement this process every day across all classrooms. This process is also inherently coupled with the KDIs for children's language and approaches to learning, where children make plans, express them verbally, follow through on their intentions, focus on activities that interest them, and reflect on their experiences.

Across all the Head Start classrooms observed, the planning time always came five minutes before work time. The lead teachers and the assistant teachers would break up into their small groups, with about 6-7 children in each sitting around the table and discuss what they were planning to do during their free play. The teacher would give children some materials to work with while they sat down with their small group and go one-by-one to each child to have them show where they would work. For example, one teacher I observed gave children a piece of paper to draw out the area that they would go play at. Another teacher used paper cameras for children to take a picture of where they would play. The lead teacher and assistant teacher would have different planning time activities for children, as noted by this observation in my field notes:

> "To plan what they will do for work time, Ruth, the lead teacher, has the children tell her where they want to play and has them point to the letter chart for the letter that stands for that area (for example, B for Block area). Then, she asks them what they will do there and if anyone else will join them. Gisele, the assistant teacher, uses two small blocks as imaginary cell phones and she gives one block/phone to the child and she has the other

> phone. They both hold up the block to their ear. She calls the child via phone making a 'ring, ring' sound and asks what they will do at work time."

In addition to addressing the language and approaches to learning KDIs that are built into this process already, teachers more often than not incorporated additional KDIs in the planning activities. In the example above of planning time, the lead teacher included a letter identification KDI in the activity for children to show her the first letter of the area where they will go play. The assistant teacher touched on a science KDI where children explored and used different technology. These various activities during the planning time do not necessarily need to be related to children's proposed activity, although in some instances they could be. For example, the teacher introduced a different way for the children to use the blocks so that gave them the idea of using the blocks as a cell phone during their work time to talk their friend and plan a birthday party. I observed very few instances in which there was an innovative planning activity was connected with the subsequent work activity.

After identifying where they would go play, teachers would often ask the children what they planned to do when they got there and how they would do it. In some instances when teachers would be pressed for time, they would simply sit down with their small group and ask children to verbally plan out what they would do without any accompanying task. High Scope does not require that teachers always have a concrete activity to complete when planning what to do for work time. Teachers also noted that High Scope encourages the creative use of materials in the classroom, and planning time is one of many instances during the day where they can incorporate different materials for children to interact with.

The 'do' component of plan-do-review is typically known as work time, or free play time. This 50-minute part of the daily routine encourages children to carry out their plan and to play with purpose. In all classrooms, I noticed that the lead teacher and the assistant teacher

closely observed children's play and remembered their plan. Whenever children would stray from their plan, the teachers would gently remind children what they said they would do but children were not required to return to their plan in any way if they changed their mind. In one classroom, a teacher was in the kitchen area playing restaurant with another group of children when she noticed from afar that one child, who had planned to work in the art area to make a collage, ended up moving to the block area to play with other children. The child left the materials out on the table, so she said to him "I know you were working on something – what were you doing?" The child was reminded of his plan and returned to the table, but instead of continuing to work on the collage, he cleaned up all the materials and went back to the block area. When I spoke with the teacher afterward and asked her to reflect on that interaction, she stated that she simply wanted to remind him of his plan in case he forgot about what he originally wanted to do during work time. Since work time, and the curriculum itself, are largely driven by children's interests, whether he would return to his stated plan or not was up to him. To have children reflect on their activity switching during review time, the teacher would make statements such as "I noticed you switched to another activity during work time…" to provide children with an opportunity to think and explain further their activities during recall time.

Immediately following work time and cleanup, the recall or 'review' time is the final step of the process. This is an opportunity for children to reflect on what they did during work time and what they learned, like the teacher above explained. During cleanup time in one classroom, I noticed that a teacher and assistant teacher picked up select items and placed them into a pillowcase as children were cleaning up their areas. When the teachers reconvened with their small group, they showed children this "mystery bag" of items that were used during their work time. The teachers removed the items from the bag one at a time, and the children and teacher

discussed both who used the object and how it was used. In this instance, recall time was a group activity as children were not only discussing what they did but also what they noticed other children doing during work time. Only in one classroom did I see the two teachers explicitly guide children towards make connections between the planning, working, and reviewing by asking them if they played with the friends that they said they would, if they completed the task they planned for, or why they chose a different activity than the one they set out to do. Similar to planning time, when teachers were running behind schedule they would have children verbally recall what they did without asking explicit questions to guide them in making connections across the three parts of the process.

All of the classrooms I observed implemented the plan-do-review process in that order, though the teachers all provided different activities for children to recall what they did and bring closure to work time. The planning and recalling part of the routine always came before and after children's free play during work time and this was observed every day across all classrooms. This plan-do-review process was part of the daily classroom routine and clearly outlined in the daily schedule (as shown in Appendix Table 3.1), so this component of the curriculum was consistently implemented across all classrooms but to varying degrees (i.e., different activities to plan and recall that targeted different KDIs).

One important thing to note about the plan-do-review process across all four classrooms was that while the children did make a plan, carry it out, and review what they did, there was never an instance where teachers asked children what they learned. When I followed up on this with the teachers after the observations, all of them said that their goal during this time to recall with children was to help them put their ideas and experiences into words in order to facilitate their language development. Children learned to express through language what they did and

how it was done but they were never explicitly asked about what exactly it was that they learned. One teacher specifically told me that this process also helps children's higher-order cognitive skills. This teacher asked children draw out what they did during work time and explain the picture to other children in the small group by using sequential language (where they played first, next, and last). High Scope would agree with the teacher that this specific activity she chose to implement also promotes children's temporal order awareness, and not just a language activity.

Child-Centered Active Learning

In an active learning setting for High Scope, children choose activities and materials that interest them, manipulate materials in their own ways, use language to describe their intentions and actions, and receive teacher support during their play. The teachers integrate the different KDIs naturally when it fits in with children's play. Active learning matches what children want to do (their interests) with what they need to do (the KDIs). Across all the classrooms I observed, active learning happened most often during work time. There were abundant materials that children could use in many ways. Although the materials were not identical in the four classrooms, they were all real, practical, everyday objects that allowed children to explore and understand the functional use of these objects. For example, during work time children engaged in dramatic play the most with all of the teachers and used a variety of materials:

> "As Teacher Carola walks around the classroom, another child is wearing a white lab coat, pretending to be a doctor and has a stethoscope. Carola asks the child who she is and the child says she is a doctor so Carola pretends to be coughing as a sick patient of the doctor. The teacher asks the child why she is sick and the child says she has a cough and with some medicine she should be better. The child hands Carola an empty container of breath mints and says that if Carola takes this medicine for a few days she should be better. Another child joins in with the play and says he has the chickenpox and the child playing doctor moves onto the next patient. Carola thanks the child for the medicine and gets up to check in on other children in the classroom."

Teachers across all classrooms supported children by entering their play, matching their developmental level and affect. Even though the dramatic play was driven by the child's interest in playing doctor, the teacher was able to address multiple KDIs. They took on roles and scenarios that children assigned to them based on the child's own experience going to the doctor (the child is a doctor, the teacher plays a sick patient, and each demonstrate an understanding of their role) and asked clarifying questions to support their language and communication (the teacher asked the child what the diagnosis was), but they also made sure to not lead the play (the child moved onto another patient and the teacher did not press any further). The teacher promoted children's dramatic play by providing real items that could be used in novel ways based on their interest (the container of breath mints as a box of medicine).

Teachers also made conscious decisions about introducing different materials into children's play throughout the week if they noticed that children had interest. For example, in one classroom a child wanted to make a paper airplane so the assistant teacher showed him how to fold a piece of paper into one during work time. Eventually, a few other children became interested and joined in. The lead teacher noticed that they were interested in making paper airplanes so she brought out some plastic model airplanes for children to play with during outside time. Children ran around outside pretending to fly the planes and some brought their paper airplanes out to fly, practicing their motor skills (a physical development KDI). The children wanted to take the model airplanes home but the teacher explained to them that they had to leave it in the classroom to share with their friends from the morning session. The teacher reassured the children that they could play with them again the following day during work time. In this instance, the teacher provided a variety of materials that lent themselves to children's interest and exploration.

Across all four classrooms, I found that teachers implemented child-centered active learning true to the High Scope curriculum. The teachers all connected children's interests with what children needed to learn via the KDIs. They entered children's play with the intent of enhancing specific KDIs when they related to activities in progress. Teachers asked how many cards children won in a game of memory, whether a block structure is taller or shorter than the children who built it, what groceries a child has scribbled on a pretend grocery list, and what children predict about whether an object will sink or float. The assumption here is that such interactions built on children's interests will support their play and learning.

Additionally, a wealth of materials was provided for children to interact with and engage in with other children, teachers thoughtfully added in new materials that reflected children's interests, and the classroom was organized in way that accommodated a wide range of activities and different types of play. The High Scope manual also provides comprehensive lists of materials to incorporate into the classroom and during play (Epstein, 2014, pp. 42-44) and encourages teachers to recognize what children are interested in, which might have helped teachers implement this key component of High Scope regularly across the different classrooms.

Learning Across all Domains

High Scope's KDIs are statements related to children's development that provide a framework for active learning. KDIs are incorporated throughout every part of the High Scope day but varied in their complexity. All of the teachers implemented this component of the High Scope curriculum by explicitly writing in their lesson plans which KDIs would be addressed throughout the week. At this particular Head Start agency, teachers are provided with the same lesson plan template to fill out (shown in Appendix Figures 3.1 and 3.2) in which teachers are required to address at least one KDI that they will focus on for the week across the eight content

domains. At each observation site, I recorded the KDIs that teachers planned for the week and received copies of their lesson plans. I observed that while teachers were able to implement learning across all domains in their classroom activities with the KDIs as a guide, each classroom had different KDIs and corresponding activities. Teachers noted that this was because they planned their lessons out depending on the level that children were at in their classroom, their interests, and if any special events were happening that week either in the class or at the center. Table 3.2 displays an example from one classroom of the KDIs that covered the week I observed. Because High Scope is a whole-child curriculum, the lesson plans throughout the week were wide in scope to address all domains of children's learning development. Below, I give a few examples of some of the domains and their corresponding KDIs from Table 3.2.

[Insert Table 3.2]

During work time, the teacher gave children the option of participating in a dot art if they had an interest in it. If they did not, then they were free to work on what their original plan was. In this instance the teacher promoted children's engagement by giving them the choice to focus on an activity that interested them (approaches to learning). To promote their physical development, during large-group time the teacher danced with children to the Hokey Pokey song where they listened to the directions of the song to put their right hand in, right hand out, and shake it all about. Here, the teacher promoted children's body awareness by having them physically navigate their body in space (perceptual, motor, and physical development). As an example of a more specific academic content area, the teacher addressed children's alphabetic knowledge during planning time. The teacher gave each child their name card and one by one, asked them to identify and sound out the first letter of their name. As a group, the teacher and

children put the name cards in alphabetical order to decide who will get to say their plan out loud first (literacy).

After my observations, I asked teachers about how they decide what KDIs to address in their lesson plans. One teacher reported that High Scope does not explicitly provide a sequence or developmental progression to follow in promoting learning across all domains, but rather provides a flexible system for lesson planning. The teachers all reported that they are able to implement this part of the curriculum consistently because their Head Start agency requires that they address at least one KDI in each content domain weekly. This is actually a specific requirement of this agency, but not of all Head Start programs or High Scope itself. On the lesson plan template, the teachers are required to write down the specific KDI that corresponds with the classroom activities planned. More than one KDI can be addressed in the activity and there can be multiple activities for each KDI. Upon further inspection, the High Scope manual does provide a list of strategies and suggestions for each of the content areas of the KDIs but it is up to the teacher to plan the KDIs according to what they think their class needs.

Scaffolding Children's Learning

To implement High Scope's conceptualization of child-centered active learning and learning across all domains requires teacher scaffolding. In High Scope, the role of the teacher is to support and extend the children's learning by observing and listening, asking appropriate questions, and scaffolding learning experiences. Across the four classrooms, I observed that teachers scaffolded children's learning at a rudimentary level. For example, teachers would ask open- and closed-ended questions such as "What is your robot's head made out of? What shape is that?" "If we mix the blue and yellow paint together, what color do we make?" "How can we

make this building as tall as Ivan?" In one of my observations I noted this scenario during work time in the classroom:

> "Teacher Edith is in the block area with a group of children and she asks them about the structure they're building. The children explain that they are building a castle and need rectangular and square blocks for the walls. She asks them where the door to the castle is, how many rooms there will be, and how many levels there are to the castle. As the children are putting up the walls to the castle, she asks them to identify the shapes – rectangle, square, triangle. Edith asks, 'Jason, what shape are you holding?' Jason says it is a square and he is using it as the door of the castle. She asks him how a square is different from a triangle, and he said they have a different number of sides. Then, Edith turns to the other children who are working with Jason and asks them similar questions. By the time Edith turns back around to Jason to see where he is at in his building, Jason has moved onto making a tower by stacking blocks on top of each other. Edith asks Jason what happened to his castle and he said that he didn't want to build it anymore because he ran out of squares and rectangles to put together the rest of the walls."

In this particular scenario, the teacher listened to Jason as he told her what he was building and when the time was right, she asked him to identify the different shapes and how the shapes were different from one another. Even during this free play time when both the adult and child are engaged in active participatory learning, the teacher addressed a KDI from math, in which she asked Jason to identify, name, describe, and compare different shapes. Since Jason demonstrated a good understanding of basic shapes, perhaps the teacher could have probed his thinking a little further and asked him how a square and rectangle, which both have the same number of sides, are different from one another. In the instance where Jason turned to a different structure because he ran out of walls, the teacher could have extended his thinking by asking him to consider what other ways he could make his wall (i.e., putting together two triangles to form a square).

While scaffolding happened most often during children's work time, it is meant to happen throughout the day for other activities as well. For instance, during small group time one teacher had children searching for miniature dinosaurs she had placed around the classroom. She asked them to collect the dinosaurs they saw and to bring it back so they could each count them.

As children brought back their dinosaurs, some counted as many as ten, and then she asked them to sort their dinosaurs according to their color and to identify the colors. For the children who were more advanced in their counting and sorting, in an effort to scaffold their learning a bit further the teacher asked individual children to count the number of dinosaurs within each color group or to add them up with another child's dinosaurs and count how many. Even though the teacher provided a challenge for some children, these scaffolding strategies were not inferential in nature (i.e., did not encourage children's higher order understanding about the lesson). High Scope might suggest that teachers have children combine all of the dinosaurs together and make a pictograph to represent the number of different colored dinosaurs as a way to address additional KDIs other than simply counting.

In theory, the High Scope curriculum allows for more dynamic processes of implementation, such as differentiation of lessons and engaging in high-quality conversations, because it is not a scripted curriculum that teachers have to adhere to. High Scope suggests that teachers ask clarifying questions to scaffold children's thinking: "to support the children, the teacher may ask what else the child did, restate what the child said and add more detail, and add context to what the child said. To offer an extension, the teacher may want to ask the child what they did with the materials, ask what else the child remembers about the experience, or encourage the child to recall more activities" (High Scope, 2013). High Scope's suggestions about scaffolding are to ask open-ended questions, but across all classrooms, I observed closed-ended scaffolding questions more than 50% of the time throughout the day. Because scaffolding is not clearly modeled or made explicit in the curriculum, these complexities may contribute to my observations of similar high support scaffolding strategies across classrooms (i.e., high

amount of guidance by asking children to produce a correct answer to a task; Justice & McGinty, 2009).

When I asked teachers about their scaffolding strategies in the classroom, they loosely expressed that they tried to scaffold by looking for "natural opportunities" to engage with children in language and communication. For example, the vignette displayed in Table 3.1 for scaffolding children's learning was a rich interaction I observed in the classroom where the teacher incorporated multiple KDIs across different learning domains including science, language, and mathematics to scaffold their understanding of why some items sink and some items float in water. The teacher built on children's prior understanding of what it means for an item to float or sink by reinforcing what they already know, she asked children to hypothesize why some items might sink or float (incorporating a science KDI), and when children got stuck, she reframed her question so that they could pick up the items themselves and compare their weights (incorporating a math KDI), and introduced them to a new word to describe the *experiment* they were conducting (incorporating a literacy KDI). In this particular instance, not only did the teacher scaffold children's learning, but she also promoted learning across multiple domains and engaged in active learning by focusing on the children's interest with water. Overall, my observations of teachers' scaffolding suggest that they do in fact implement this curriculum component, but to varying levels of support and scaffolding questions across classrooms.

Discussion

The present study conducted a qualitative evaluation of what goes on inside classrooms implementing a whole-child curriculum through observations of fidelity in four local Head Start classrooms. I focused my observations on the High Scope curriculum, which was originally

conceived to be a flexible curriculum tailored to children's rapidly changing interests and needs, but little is known about how it is actually tailored or deviated from the intended curriculum. At the surface level, my findings revealed that teachers "checked the boxes" of doing what the High Scope curriculum loosely prescribes. Teachers all integrated High Scope's curriculum content—the KDIs—throughout the Head Start day for children, and language development and support played a big role across the classrooms. Teachers engaged with children in the plan-do-review process that was prescribed as part of the daily routine, promoted child-centered active learning by setting up a classroom environment where children were free to explore based on their interests and supporting their play, incorporated learning across all domains with the guidance of the KDIs in their lesson planning, and finally, they scaffolded children's thinking by asking them questions when the opportunities presented themselves naturally.

However, I found that teachers varied in their implementation of the more nuanced aspects within each of the components of the curriculum. For High Scope's plan-do-review process, the sequence of activities is meant for children to plan, enact, and reflect on their play so they can communicate their actions and create mental representations of what they did to support their executive function. Three of the four classrooms I observed deviated away from this in that the planning and reflecting of children's work time was mostly to support children's language. Only one classroom I observed was intentional about both supporting children's language development and their higher-order cognitive skills by having them explicitly reflect sequentially what they did first, second, and last during their free play time with the help of picture drawings. To promote child-centered active learning, teachers provided abundant materials and activities for children to choose from and made conscious decisions regarding what materials should be incorporated into the activities. As children played and explored around the classroom, the

teachers looked for natural opportunities to probe children's thinking and incorporate learning across various domains of their school readiness. While teachers addressed different domains in their small- and large-group activities—touching on literacy, science, approaches to learning—the level of instructional support teachers demonstrated when teaching was rather low from a purely observational standpoint. Similar to Odom et al. (2010), my own logic suggests that *full* implementation in early childhood curriculum research rarely occurs.

For ideal High Scope implementation, to encourage active learning across all domains relies heavily on teacher scaffolding. The majority of my classroom observations revealed that the level of scaffolding teachers provided to children were at a most rudimentary level across all classrooms. At a minimum, teachers mostly asked closed-ended questions and provided a high amount of guidance by asking children to produce a correct answer to a task. For example, one teacher I observed said, "I see you're playing with some shapes. What shapes are those? How many sides does it have? How many corners does it have?" I observed multiple opportunities for teachers to take advantage of an unexpected student comment to build content knowledge and promote children's higher-order thinking consistent with High Scope (e.g., "What makes this shape a rectangle? What makes this a circle? How do we know this is a shape?"), but teachers adapted the curriculum into basic questioning and relied heavily on knowledge of automatized responses. Under ideal conditions, implementing the High Scope curriculum as intended could potentially support teachers in moving toward a different kind of instructional discourse during formal small-group lessons and informal teacher-child interactions and activities throughout other parts of the day.

Learning to embed the High Scope KDIs and integrate across all different domains of school readiness by guiding individual children's development and scaffolding their learning

within the various High Scope activities throughout the day is a formidable task yet is expected of teachers in this whole-child, constructivist tradition. Scaffolding is arguably the most essential component of any early childhood curriculum (Bodrova & Leong, 2007; Berk & Winsler, 1995), but it also the most difficult out of all the High Scope curriculum components to implement because it is meant to be flexible and responsive to individual children and also requires teachers to utilize a variety of strategies. This takes considerable skill on the part of the teacher. Indeed, Hamre, Downer, Jamil, and Pianta (2012) note that teachers' capacities for enacting effective interactions with children are dependent in part on their skills in seeing the moment-to-moment cues and features in the flow of interaction as it unfolds in real time. My finding that teachers struggle to scaffold children's play and thinking at a more complex level is also consistent with Chien et al.'s (2010) findings from a larger scale study of preschool programs. Their study revealed that levels of teachers' stimulation of language and conversations, feedback and scaffolding of skills, and focus on conceptual understanding were rather low. Of note is that preschool programs in general struggle most with instructional support and the classrooms I observed were no exception. In fact, the Head Start umbrella agency of these classrooms conducts routine observations of the *Classroom Assessment Scoring System* (CLASS; Pianta, La Paro, & Hamre, 2008) and found that to be the case as well for the instructional support domain of the CLASS measure.

High Scope allows for a great deal of teacher choice in the curriculum as constructivist theories respect teachers as decision-makers and expect them to design daily learning activities and interactions to meet individual and classroom needs and interests. However, there is also a strong need for the curriculum to provide a framework for making these decisions, as well as specific and clear methods, directions, examples, and sequences for achieving learning

objectives. My close examination of the High Scope curriculum suggests that there is currently not one in place. There must be a delicate balance between specifying a scope and sequence of activities and instructional procedures that describe expected teacher behavior and practices, but also opportunities for teachers to exercise their own independent judgment and differentiate their teaching practices based on children's ongoing needs. Clements and Sarama (2008) provide one example of such an approach with their Building Blocks mathematics curriculum in which specific instructional behaviors are clearly delineated for teachers when they deliver the curriculum. Even though this is skill-specific curricula, it could potentially be adapted to align with a whole-child curriculum. Further, the current study only focused on one of many whole-child curricula that are being implemented in preschool programs across the country today. My study was limited in that I only evaluated the High Scope curriculum, but it will be important to validate my observations with other studies of classrooms such as those implementing Creative Curriculum. One important question to answer with future research is whether whole-child curricula are qualitatively different from one another in their implementation and adaptation across different preschool settings.

Implications and Directions for Future Research

Measurement of implementation fidelity for curricular programs and interventions is a fruitful area for future studies as it can help researchers explain why an intervention did or did not produce the expected results. Implementation science is emerging as a new and important dimension of early childhood research (Mendive et al., 2016; Warren, Domitrovich, & Greenberg, 2009), and the questions that were once unacknowledged or viewed as subtle features of studies (e.g., the manner in which researchers assesses implementation) are emerging as prominent factors in curriculum evaluation research. Further, there is a considerable need for

more research on factors that enhance or relate to adequate implementation of curricular programs and lead to effective classroom practices and child outcomes. My study contributes to the emerging literature on measurement of implementation in early childhood curriculum research by demonstrating the utility of a novel approach to assessing fidelity of implementation. The results from this theory-driven qualitative evaluation of curriculum implementation might very well have been different if observers came in with checklists and rated teachers on whether the components of the curriculum were present or absent in their instruction. The purpose of theory-driven evaluation is not only to assess whether a program works or does not work, but also how and why it does so. The information is essential for stakeholders and decision-makers to improve their existing or future programs (Chen, 2005).

Specifically, this work has a number of implications in thinking about what sustainable teacher supports and development opportunities are needed in order to implement a whole-child curriculum faithfully. Use of whole-child curricula will require additional resources, as a curriculum can only be implemented well with ongoing training and professional development. There is some promising evidence of professional development models being scaled up (Downer et al., 2011; Pianta, Mashburn, Hamre, & Justice, 2008). Teacher professional development and coaching are thought to be key to educational reform (Sarama, DiBiase, Clements, & Spitler, 2004), with increases in teacher professional development and in-service education linked to improvements in classroom quality and children's development (Bowman, Donovan, & Burns, 2001; Neuman & Cunningham, 2009; Pianta, Barnett, Burchinal, & Thorrnburg, 2009; Powell, Diamond, Burchinal, & Koehler, 2010). Supports for teachers may come as part of the activity prescribed by the curricula itself, such as the use of scripting which embeds these types of instructional conversations into the activity plans. However, it will be important to make sure

that the curriculum supports remain flexible enough that teachers can include unplanned and spontaneous learning. Therefore, curriculum developers and preschool programs will both need to attend to ways that support teachers' use of these practices.

There is now ample research suggesting that such supports are effective if they are fairly intensive and ongoing (Caswell & He, 2008; Dickinson & Caswell, 2007; Domitrovich et al., 2009; Landry, Swank, Smith, Assel, & Gunnewig, 2006; Pence et al., 2008), such as monthly and summer workshops, access to online supports for the curriculum, personalized classroom observation and coaching, and video-based observations and time for reflection and discussion with coaches (Clements & Sarama, 2008; Kinzie et al., 2014). There also may be a need to train coaches as well in order to effectively help teachers implement the key features of curricula in order to increase the ultimate effectiveness of these programs. To be effective, these training models need to be provided by individuals who are also familiar with adult learning principles and the realities of teachers' classrooms. Preschool programs must look for comprehensive professional development and make sure that there is research on its effectiveness. These decision-makers must also examine the types of built-in supports available for providing ongoing technical assistance, including those at the site level such as supervisors, directors, or coaches. Moreover, they must also determine the supports that can be available within their organizational context, such as financial resources and time for teacher professional development.

It is important to note that I observed these classrooms within the context of Head Start and their policies and regulations. In addition to implementing the High Scope curriculum, the teachers also had to do what the federal Head Start program required of them. Upon re-reading my observational field notes, it emerged that teachers often wanted to share that their implementation practices were often nestled within a larger policy context and it became clear

that teachers often conflated the CLASS with High Scope curriculum implementation. Head Start programs are mandated to use the prescribed curriculum as a framework to guide interactions and activities in the classroom but must also meet the demands of the CLASS, a high-stakes measure that influences program planning, evaluation, funding, and accountability. This is another limitation to my study as it became a post-hoc finding after I deductively coded the data. Thus, I was not able to explore this closer but future work should examine the extent to which dimensions of the CLASS overlap with any particular curriculum, whole-child or not. My own future work will examine the degree to which implementation fidelity is associated with classroom quality. If so, then this will have policy implications for not putting so much pressure on teachers to direct their instruction and classroom practices on a high-stakes measure but rather focus on implementing a curriculum as faithfully as possible. Classroom "quality" in this case is measured by the fidelity of implementation of the curriculum. To be sure, this will require good and careful measurement of fidelity.

My work also speaks to the literature on understanding the role of teachers as "street-level" policymakers (Lipsky, 1980) in implementing curricula. Lipsky viewed teachers as street-level, or "on the ground," professionals whose position requires them to make sensitive observations and judgments in immediate, direct, and personal encounters with students. Because teachers enact public policy—in the case of this study, enacting a whole-child curriculum as mandated by federal Head Start—within the specific, particular, local context of interactions with individual children, each teacher-child interaction represents a unique classroom policy encounter. Research focused on the curricular and instructional experiences of teachers in various public preschool settings, such as Head Start and state-funded preschool, who have to

enforce federal and state policies would be very useful in understanding how and the degree to which a curriculum is implemented with fidelity.

In sum, programs and researchers are challenged to bridge the gap between efficacy trials and "real world" classrooms. My study aimed to understand the process and conditions by which a widely used curriculum is implemented in order to help move preschool programs forward towards greater benefits for children. Understanding the idiosyncrasies and complexities of whole-child curriculum implementation in large preschool programs operating at scale may well inform the movement of research findings into practice in early childhood classrooms.

References

Al Otaiba, S., Conner, C., Lane, H., Kosanovich, M. L., Schatschneider, C., Dyrlund, A. K., et al. (2008). Reading First kindergarten classroom instruction and students' growth in phonological awareness and letter naming-decoding fluency. *Journal of School Psychology, 46,* 281–314.

Berk, L., & Winsler, A. (1995). 'Vygotsky: His life and works' and 'Vygotsky's approach to development'. In *Scaffolding children's learning: Vygotsky and early childhood learning* (pp. 25–34). Washington, DC: National Association for the Education of Young Children.

Belfield, C. R., Nores, M., Barnett, W. S., & Schweinhart, L. J. (2006). The High/Scope Perry preschool program. *Journal of Human Resources, XLI*(1), 162-190.

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., et al. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 79,* 1802–1817.

Bodrova, E., & Leong, D.J. (2007). *Tools of the Mind: The Vygotskian approach to early childhood education* (2nd ed.). Upper Saddle River, NJ: Pearson Merrill/Prentice Hall.

Caswell, L., & He, Y. (2008). *Promoting children's early language and literacy development in contexts of early educational intervention and care: A review of the impact of federally funded research initiatives on young children's school readiness.* Washington, DC: US Department of Health and Human Services.

Chen, H. T. (1990). *Theory-driven evaluations.* Thousand Oaks, CA: Sage Publications.

Chen, H. T. (1997). Applying mixed methods under the framework of theory-driven evaluations. *New Directions for Evaluation*, *1997*(74), 61-72.

Chen, H. T. (2005). Theory-driven evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 415-419). Thousand Oaks, CA: Sage Publications.

Chen, H. T. (2013). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective.* Thousand Oaks, CA: Sage Publications.

Chen, H.T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*(3), 283-301.

Chen, H.T., & Rossi, P. H. (1992). *Using Theory to Improve Policy and Program Evaluations*. Santa Barbara, CA: Greenwood Press.

Chien, N. C., Howes, C., Burchinal, M., Pianta, R. C., Ritchie, S., Bryant, D. M., ... & Barbarin, O. A. (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development*, *81*(5), 1534-1549.

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education*, 35-70.

Clements, D., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*(2), 443-494. doi: 10.3102/0002831207312908

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, *39*(3), 124-130.

Dane, A., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18,* 23– 45. doi: 10.1016/S0272-7358(97)00043-3

DeVries, R., & Kohlberg, L. (1987). *Programs of early education: The constructivist view.* White Plains, NY: Longman.

Dickinson, D. K., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly*, *22*(2), 243-260.

Domitrovich, C.E., Gest, S. D., Gill, S., Bierman, K.L., Welsh, J.A., & Jones, D. (2009). Fostering high-quality teaching with an enriched curriculum and professional development support: The Head Start REDI program. *American Educational Research Journal*, *46*(2), 567-597.

Downer, J. T., Pianta, R.C., Fan, X., Hamre, B.K., Mashburn, A., & Justice, L. (2011). Effects of web-mediated teacher professional development on the language and literacy skills of children enrolled in prekindergarten programs. *NHSA Dialog, 14*(4), 189–212. doi:10.1080/15240754.2011.613129

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107-115.

Epstein, A. (2014) *Essentials of Active Learning in Preschool*. Ypsilanti, MI: High/Scope Press.

Epstein, A. S. (2016). *Key Developmental Indicator (KDI) scaffolding charts: The High/Scope preschool curriculum.* Ypsilanti, MI: High/Scope Press.

Epstein, A. S., & Hohmann, M. (2012). *The High/Scope preschool curriculum.* Ypsilanti, MI: High/Scope Press.

Epstein, A. S., Schweinhart, L. J., & McAdoo, L. (1996). *Models of early childhood education.* Ypsilanti, MI: High/Scope Press.

French, G. (2012). The High Scope Approach to Early Learning. In M.M. Mahuna & M. Taylor (Eds.), *Early childhood education and care: an introduction for students in Ireland* (pp. 127-134). Dublin: Gill and McMillan.

Greene, J. C. (2015). The emergence of mixing methods in the field of evaluation. *Qualitative Health Research*, *25*(6), 746-750.

Hamre, B. K., Downer, J. T., Jamil, F. M., & Pianta, R. C. (2012). Enhancing teachers' intentional use of effective interactions with children: Designing and testing professional development interventions. In R.C. Pianta (Ed.), *Handbook of early childhood education* (pp. 507–532). New York, NY: Guilford Press.

Heckman, J. J., Smith, J., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, *64*(4), 487-535.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288.

Jenkins, J. M., & Duncan, G. J. (2017). Do pre-kindergarten curricula matter? In D. Phillips, K. A. Dodge, & Pre-Kindergarten Task Force (Eds.), *The current state of scientific knowledge on pre-kindergarten effects* (pp. 37-44). Washington, DC: Brookings Institution and Duke University.

Jenkins, J.M., Duncan, G.J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. (2018). Boosting school readiness: Should preschools target skills or the whole child? *Economics of Education Review*. doi: 10.1016/j.econedurev.2018.05.001

Kinzie, M. B., Whittaker, J. V., Williford, A. P., DeCoster, J., McGuire, P., Lee, Y., & Kilday, C. R. (2014). MyTeachingPartner-Math/Science pre-kindergarten curricula and teacher supports: Associations with children's mathematics and science learning. *Early Childhood Research Quarterly*, *29*(4), 586-599.

Krueger, A. B., & Zhu, P. (2004). Another look at the New York City school voucher experiment. *American Behavioral Scientist*, *47*(5), 658-698.

Landry, S. H., Swank, P. R., Smith, K. E., Assel, M. A., & Gunnewig, S. B. (2006). Enhancing early literacy skills for preschool children bringing a professional development model to scale. *Journal of Learning Disabilities*, *39*(4), 306-324.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Newbury Park, CA: Sage Publications.

Lipsky, M. (1980). *Street level bureaucracy: Dilemmas of the individual in public services.* New York, NY: Russell Sage Foundation.

Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program. *Journal of Educational Psychology*, *108*(1), 130-145.

Moiduddin, E., Aikens, N., Tarullo, L., West, J., & Xue, Y. (2012). *Child outcomes and classroom quality in FACES 2009.* OPRE Report 2012-37a. Washington, DC: Administration for Children & Families.

Nguyen, T. (2017). Impacts of targeted and global preschool curricula on children's school readiness: A meta-analytic review. Working Paper, Irvine Network on Interventions in Development.

Odom, S. L. (2009). The tie that binds: Evidence-based practice, implementation science, and early intervention. *Topics in Early Childhood Special Education, 29,* 53–61.

Odom, S. L., Fleming, K., Diamond, K., Lieber, J., Hanson, M., Butera, G., ... & Marquis, J. (2010). Examining different forms of implementation and in early childhood curriculum research. *Early Childhood Research Quarterly*, *25*(3), 314-328.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*, 33–84.

Patton, M.Q. (2014). Qualitative inquiry in utilization-focused evaluation. In L. Goodyear, E. Barela, J. Jewiss, & J. Usinger (Eds.), *Qualitative inquiry in evaluation: From theory to practice*. San Francisco: Jossey-Bass.

Pence, K. L., Justice, L. M., & Wiggins, A. K. (2008). Preschool teachers' fidelity in implementing a comprehensive language-rich curriculum. *Language, Speech, and Hearing Services in Schools, 39,* 329–331.

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

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of Web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*(4), 431–451. doi:10.1016/j.ecresq.2008.02.001

Ramey, C. T., Bryant, D. M.,Wasik, B. H., Sparling, G. J., Fendt, K. H., & La Vange, L. M. (1992). Infant Health and Development Program for low birth weight, premature infants: Program elements, family participation, and child intelligence. *Pediatrics, 89,* 454-465.

Sarama, J., DiBiase, A. M., Clements, D. H., & Spitler, M. E. (2004). The professional development challenge in preschool mathematics. *Engaging young children in mathematics: Standards for early childhood mathematics education*, 415-446.

Schweinhart, L. J. (2005*). Lifetime Effects: The High/Scope Perry Preschool Study through Age 40.* Ypsilanti, MI: High/Scope Educational Research Foundation.

U.S. Department of Health and Human Services. (2016). *Head Start Program Performance Standards.* Retrieved from https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/hspps-appendix.pdf.

Vygotsky, L. S. (1978). Interaction between learning and development. *Readings on the Development of Children, 23*(3), 34-41.

Warren, H. K., Domitrovich, C. E., & Greenberg, M. T. (2009). Implementation quality in school-based research: roles for the prevention researcher. In L. M. Dinella (Ed.), *Conducting science-based psychology research in schools* (pp. 129-151).Washington DC: American Psychological Association

Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New directions for evaluation*, *1997*(76), 41-55.

Wiltshire, M. (2012). *Understanding the High/Scope approach: Early years education in practice.* London, UK: Routledge.

Zigler, E. F., & Bishop-Josef, S. J. (2006). The cognitive child versus the whole child: Lessons from 40 years of Head Start. *Play=learning: How play motivates and enhances children's cognitive and social-emotional growth*, 15-35.

Figure 3.1. High Scope Key Developmental Indicators for learning across all domains.

HighScope Educational Research Foundation

HighScope Preschool Curriculum

Key Developmental Indicators (KDIs)

A. Approaches to Learning

1. **Initiative:** Children demonstrate initiative as they explore their world.
2. **Planning:** Children make plans and follow through on their intentions.
3. **Engagement:** Children focus on activities that interest them.
4. **Problem solving:** Children solve problems encountered in play.
5. **Use of resources:** Children gather information and formulate ideas about their world.
6. **Reflection:** Children reflect on their experiences.

B. Social and Emotional Development

7. **Self-identity:** Children have a positive self-identity.
8. **Sense of competence:** Children feel they are competent.
9. **Emotions:** Children recognize, label, and regulate their feelings.
10. **Empathy:** Children demonstrate empathy toward others.
11. **Community:** Children participate in the community of the classroom.
12. **Building relationships:** Children build relationships with other children and adults.
13. **Cooperative play:** Children engage in cooperative play.
14. **Moral development:** Children develop an internal sense of right and wrong.
15. **Conflict resolution:** Children resolve social conflicts.

C. Physical Development and Health

16. **Gross-motor skills:** Children demonstrate strength, flexibility, balance, and timing in using their large muscles.
17. **Fine-motor skills:** Children demonstrate dexterity and hand-eye coordination in using their small muscles.
18. **Body awareness:** Children know about their bodies and how to navigate them in space.
19. **Personal care:** Children carry out personal care routines on their own.
20. **Healthy behavior:** Children engage in healthy practices.

D. Language, Literacy, and Communication[1]

21. **Comprehension:** Children understand language.
22. **Speaking:** Children express themselves using language.
23. **Vocabulary:** Children understand and use a variety of words and phrases.
24. **Phonological awareness:** Children identify distinct sounds in spoken language.
25. **Alphabetic knowledge:** Children identify letter names and their sounds.
26. **Reading:** Children read for pleasure and information.
27. **Concepts about print:** Children demonstrate knowledge about environmental print.
28. **Book knowledge:** Children demonstrate knowledge about books.
29. **Writing:** Children write for many different purposes.
30. **ELL/Dual language acquisition:** (If applicable) Children use English and their home language(s) (including sign language).

[1]KDIs 21–29 may be used for the child's home language(s) as well as English. KDI 30 refers specifically to ELL/dual language acquisition.

E. Mathematics

31. **Number words and symbols:** Children recognize and use number words and symbols.
32. **Counting:** Children count things.
33. **Part-whole relationships:** Children combine and separate quantities of objects.
34. **Shapes:** Children identify, name, and describe shapes.
35. **Spatial awareness:** Children recognize spatial relationships among people and objects.
36. **Measuring:** Children measure to describe, compare, and order things.
37. **Unit:** Children understand and use the concept of unit.
38. **Patterns:** Children identify, describe, copy, complete, and create patterns.
39. **Data analysis:** Children use information about quantity to draw conclusions, make decisions, and solve problems.

F. Creative Arts

40. **Art:** Children express and represent what they observe, think, imagine, and feel through two- and three-dimensional art.
41. **Music:** Children express and represent what they observe, think, imagine, and feel through music.
42. **Movement:** Children express and represent what they observe, think, imagine, and feel through movement.
43. **Pretend play:** Children express and represent what they observe, think, imagine, and feel through pretend play.
44. **Appreciating the arts:** Children appreciate the creative arts.

G. Science and Technology

45. **Observing:** Children observe the materials and processes in their environment.
46. **Classifying:** Children classify materials, actions, people, and events.
47. **Experimenting:** Children experiment to test their ideas.
48. **Predicting:** Children predict what they expect will happen.
49. **Drawing conclusions:** Children draw conclusions based on their experiences and observations.
50. **Communicating ideas:** Children communicate their ideas about the characteristics of things and how they work.
51. **Natural and physical world:** Children gather knowledge about the natural and physical world.
52. **Tools and technology:** Children explore and use tools and technology.

H. Social Studies

53. **Diversity:** Children understand that people have diverse characteristics, interests, and abilities.
54. **Community roles:** Children recognize that people have different roles and functions in the community.
55. **Decision making:** Children participate in making classroom decisions.
56. **Geography:** Children recognize and interpret features and locations in their environment.
57. **History:** Children understand past, present, and future.
58. **Ecology:** Children understand the importance of taking care of their environment.

Table 3.1. Components of the High Scope curriculum.

Component & Description	Example: Children are interested in water	Occurrence during the day
Plan-do-review High Scope considers this the hallmark of their curriculum; children are given an opportunity to think through and make choices about what they are going to do (planning time), implement those ideas during work time (free play), and then reflect on their activities during the review (recall) time	For planning time, Teacher Betty asks children to verbally tell her what they will do during work time. As soon as children are done planning, she asks them to think of their favorite underwater animal (e.g., fish, shark, dolphin) and swim like that animal to the area where they will work. During work time, the teacher introduces the water table and children take turns at this area. After cleanup time, the children sit back down at their small group tables and the teacher asks children to paint a picture of what they did with watercolor paint and explain to the group what they did.	Planning time; work time; recall time
Child-centered active learning Children are provided with abundant age-appropriate materials to work with; children choose materials and play partners, change and build on their play ideas, and plan activities according to their interests and needs; teachers and children talk naturally about what they are doing	Betty introduces the water table and the children decide they want to bathe the toy dinosaurs for the classroom next door as a kind gesture. Children play with water and the teacher notices how much the class is interested in it. During literacy time, instead of reading *Brown Bear*, the teacher asks children what they want to read and the class decides on a non-fiction book *A Drop of Water*. In the book, they discuss its properties and children are introduced to the word 'evaporation' and they ask what it means.	Throughout the day
Learning across all domains Teachers emphasize High Scope's eight content areas and focus on the respective Key Developmental Indicators in their lesson planning	During literacy time, Betty reads a book about water and discuss evaporation. The following day, during small-group time, children go outside and sit under the sun. They each get a sponge with a letter printed on it to dip in water and press into the concrete. Two children want the same letter, so the Betty asks how the children can resolve their conflict so that both can use the same letter sponge. The children resolve their problem and decide to take turns. The teacher reminds children about the word 'evaporation' that they learned and what the sun is doing to make their letter disappear.	Throughout the day
Scaffolding children's learning Teachers support and gently extend children's thinking and reasoning; teachers first validate, or support, what children already know, and then when the time is right, gently encourage them to extend their thinking to the next level by asking them questions	Betty noticed that the children enjoyed working at the water table during work time, so she made the water table available again for children the next day. A child, Johan, put a miniature toy turtle into the water table and said to the teacher that it was floating. She said to Johan, "Yes, the turtle is swimming on top of the water so it is floating!" Johan asked if he could put more items in the water and the teacher agreed and said he should test to see what else can float. He and the other children that joined chose a feather, a rock, a golf ball, a marble, and a magnetic tile piece. Before they put the items in the water one by one, the teacher asks children to guess what they think will float or not. As they dropped the items in the water, they noticed that some items floated while other items went all the way to the bottom. Betty explains to the children that means the item is sinking if it went all the way down to the bottom. Then she asks, "Why do you think those items were sinking while the others were floating?" The children did not answer, so she asked them differently, "If we pick up the leaf and the rock from the water and hold them in our hands, what is different about the two items? Why do you think the leaf went to the top of the water but the rock went to the bottom?" From there, Betty and the children discussed why some of the heavier items sank to the bottom. She talked about the term *experiment* and said that "When we try things out to see if our ideas are correct, this is called an experiment. How was this activity an experiment?" Johan said they did an experiment to see which items floated in the water and which ones sank.	Throughout the day

Table 3.2. Example of key developmental indicators the teacher focused on for the week.

Domain	KDI	Classroom Example	Occurrence during the day
Approaches to Learning	Engagement: Children focus on activities that interest them.	On Monday, Teacher Kristina puts out paper plates on the tables during work time and she tells the class if they want to do dot art with paint and cotton swabs, they can come over and work on that.	Work time
Social Emotional Development	Community: Children participate in the community of the classroom.	On Tuesday, Kristina tells the class that Clifford the big red dog is visiting them later this week, so they are going to decorate a house to welcome him. She tells them if they want to join, they can all work together in the art area to decorate it.	Work time
Language and Communication	Comprehension: Children understand language.	On Tuesday, Kristina reads a book about manners and discusses with the class what "please" and "thank you" mean.	Literacy time
Literacy	Alphabetic knowledge: Children identify letter names and their sounds.	During planning time on Thursday, Kristina hands out each child's name card and asks them to identify and sound out the first letter of their name. Then, as a group they decide who will get to plan first by going in alphabetical order.	Planning time
Mathematics and Development	Patterns: Children identify, describe, copy, complete, and create patterns.	During small-group time on Wednesday, Kristina reminds the children about *The Very Hungry Caterpillar* they read about. She tells them that they will make their own caterpillar with the different colored pieces of construction paper that are pre-cut into circles. She tells the group that they can make a pattern similar to the caterpillar they read about (light green, green, light green, green) or make their own pattern.	Small-group time
Scientific Reasoning	Experimenting: Children experiment to test their ideas.	During small-group time on Friday, Kristina re-introduces dot art from Monday. She gives them each three paint colors (blue, yellow, red) and asks them to paint their name onto the plate. She tells them they can mix the colors if they want to and see what they get.	Small-group time
Perceptual, Motor, and Physical Development	Body awareness: Children know about their bodies and how to navigate them in space.	Kristina plays the Hokey Pokey song for children to dance to. The children follow the directions to the song and they practice putting their different body parts into the middle (e.g., put your left foot in).	Large-group time
English Language Development	Comprehension: Children understand language.	On Friday, Kristina read a book called No Biting and the class discusses what biting means and what they can and cannot put in their mouths. She also asks them why they cannot put certain items in their mouths.	Literacy time

Appendix Table 3.1. Daily schedule for all Head Start classrooms observed.

Morning Session	Afternoon Session	Event or Activity
8:30-8:35 am	1:00-1:05 pm	Arrival and greeting time
8:35-8:45 am	1:05-1:15 pm	Large group time
8:45-9:00 am	1:15-1:30 pm	Small group time
9:00-9:05 am	1:30-1:35 pm	Planning time
9:05-9:55 am	1:35-2:25 pm	Work time
10:00-10:05 am	2:30-2:35 pm	Recall time
10:10-10:50 am	2:40-3:20 pm	Roll call and outside time
10:50-11:00 am	3:20-3:30 pm	Roll call and literacy time
11:00-11:30 am	3:30-4:00 pm	Hand washing and lunch/snack time
11:30 am	4:00 pm	Dismissal

Appendix Figure 3.1. High Scope lesson plan template for Head Start teachers, part 1.

LESSON PLAN

KEY DEVELOPMENTAL INDICATORS (KDIs): Write the Key Developmental Indicator(s) that you will focus on this week. Use at least 1 KDI in each of the 8 Domains. (Pink)	
AL Approaches to Learning	
SE Social Emotional Development	
LC Language & Communication	
LIT Literacy	
MD Mathematics & Development	
SR Scientific Reasoning	
PMPD Perceptual, Motor & Physical Development	
EL English Language Development	

DATE & LITERACY Must Add: 1) Resource 2) Focus	PLANNING / RECALL	WORK TIME	OUTSIDE TIME
Monday ☐ 1) 2)	☐	☐	☐
Tuesday ☐ 1) 2)	☐	☐	☐
Wednesdays ☐ 1) 2)	☐	☐	☐
Thursday ☐ 1) 2)	☐	☐	☐
Friday ☐ 1) 2)	☐	☐	☐

Parent Input into the Curriculum: _____

2/16

Appendix Figure 3.2. High Scope lesson plan template for Head Start teachers, part 2.

Additional Required Elements

- ☐ N Nutrition (Green)
- ☐ PI Parent Input into the Curriculum (Yellow)
- ☐ PE Physical Exercise (Daily) (Purple)
- ☐ S/PS Safety/Pedestrian Safety (done 1st week of the month) (Orange)
- ☐ H Health (Red)
- ☐ MH Mental Health Activity (Blue)

Small Group Activity Teacher One	Small Group Activity Teacher Two	Large Group
☐ Activity: Materials: Resources:	☐ Activity: Materials: Resources:	☐
☐ Activity: Materials: Resources:	☐ Activity: Materials: Resources:	☐
☐ Activity: Materials: Resources:	☐ Activity: Materials: Resources:	☐
☐ Activity: Materials: Resources:	☐ Activity: Materials: Resources:	☐
☐ Activity: Materials: Resources:	☐ Activity: Materials: Resources:	☐

CHAPTER FOUR

Study 3: Skill-Specific Preschool Curriculum Interventions: Examining Treatment Effect Heterogeneity with Distributional Methods

Preschool curriculum interventions aimed at closing achievement gaps are a rich area of research in early childhood education. Skill-specific curriculum interventions, in particular, have become increasingly popular in this past decade, stemming from an increasing focus on improving children's academic achievement as well as evidence that exposure to explicit learning opportunities may enhance the effectiveness of early childhood programs (Clements & Sarama, 2007; Hamre, Downer, Kilday, & McGuire, 2008; PCER, 2008). Much of these skill-specific curricula embed learning in playful preschool activities, including storybook reading, games, art, and discovery activities that are conducted in both small and large group contexts and grounded in a sound developmental framework. Teachers are provided with lesson plans to follow in which playful and guided activities are strategically organized to present children with learning opportunities that are intended to be focused, sequential, and cumulative.

Given the emphasis on directly targeting children's *skills* with these curricula, it thus becomes important to understand for whom these interventions are most effective along the achievement distribution. For example, children enter preschool classrooms with different skills and achievement levels that put them closer or farther from meeting goals at the end of the preschool year. As teachers work with children to meet or exceed these goals, some will need to learn more than others, while others may have already mastered the basic skills and content from the curriculum. Because children are likely to be experiencing different processes during the curriculum intervention, the purpose of the current study is to examine variation in treatment effects across the outcome distribution. I contribute to the growing literature on treatment effect heterogeneity in developmental science by applying a methodological strategy from economics

to identify which children are likely to benefit most from literacy-focused curricular interventions.

Efficacy of Skill-Specific Curriculum Interventions

Children in classrooms that implement skill-specific curricula during preschool show moderate to large improvements in the targeted content domain (Clements & Sarama, 2008; Diamond, Barnett, Thomas, & Munro, 2008; PCER, 2008; Weiland & Yoshikawa, 2013). For example, children who received a literacy-targeted curriculum intervention showed improvements in their literacy and language skills (Justice et al., 2010; Lonigan, Farver, Philips, & Clancy-Menchetti, 2011). Larger, publicly funded programs that have implemented skill-specific curricula such as Boston's universal preschool—often referred to as one of the most successful early childhood programs (Duncan & Murnane, 2014)—showed a similar pattern in findings, with children making the most gains on outcome domains that measure the specific skills targeted by the curricula (Weiland & Yoshikawa, 2013).

Few studies have compared across different types of skill-specific curricula, whole-child, and locally-developed curricula. Skill-specific curricula directly target specific school readiness domains, whereas whole-child curricula which aims to provide enriching experiences across multiple domains of children's development (e.g., health, social-emotional, academic). On the other hand, locally-developed curricula include an assortment of activities that early childhood education centers develop on their own. Two recent meta-analyses showed empirical evidence demonstrating the value of skill-specific curricula. Chambers, Cheung, and Slavin (2016) examined 22 literacy and language and whole-child programs across 32 studies published since 1999 and found evidence of positive impacts on the targeted domains at the end of preschool and kindergarten. In a meta-analysis of 71 experimental and quasi-experimental studies published

since 1990, Nguyen (2017) found small to moderate effect sizes for skill-specific literacy and language (.17 to .20 SD) and mathematics (.39 to .40 SD) curricula relative to whole-child and locally-developed curricula. Included in the calculation of the effects sizes from Nguyen (2017) are intervention impacts from the only large-scale, systematic evaluation comparing across these different types of curricula.

In the early 2000s, 12 grantees across the U.S. were funded by the Institute of Educational Sciences (IES) to study the effect of 14 preschool curricula on children's academic and social-emotional outcomes up to the end of kindergarten in the Preschool Curriculum Evaluation Research Initiative Study (PCER, 2008). The goal of the PCER study was to understand whether different widely available curricula, or specific features of these curricula, were beneficial in promoting children's learning and development during their preschool year at age four. Of the 14 intervention curricula, 10 focused on early language and literacy development, one focused on mathematics, and the other three focused on more general domains of children's development.

The findings from the PCER study were largely null, although several analytic issues, such as low statistical power (because each curriculum was evaluated individually), have been cited to explain the lack of significant effects (PCER, 2008). However, two content-specific curricula (literacy and math) significantly improved children's reading and math outcomes at the end of preschool, with improvements in the targeted domain (i.e. math curricula affecting math outcomes). The report indicates effect sizes ranging from .36 to .51 standard deviations (SD) on children's reading, language, and mathematics outcomes for the literacy curriculum DLM Early Childhood Express supplemented with Open Court Reading Pre-K and effect sizes ranging from .44 to .96 on children's math outcomes for the Pre-K Mathematics curriculum supplemented

with DLM Early Childhood Express Math software. In a reanalysis of the PCER data, Jenkins et al. (2018) pooled the skill-specific curricula together and found that compared with the High Scope and Creative Curriculum in most public pre-K classrooms, skill-specific curricula increased children's outcomes in the targeted content domain with effect sizes ranging from .15 to .35 SD. However, this reanalysis examined only simple differences in mean outcomes across the treatment and control groups for skill-specific curricula and did not consider the possibility of heterogeneous impacts.

Treatment Effect Heterogeneity

Though there is a growing evidence base on the efficacy of skill-specific curricula, it is unclear for whom this type of curricula is most effective. Given that children enter into the classroom with differing levels of skills, it is important that research discern the conditions under which certain curricula work best for which groups of children. Developmental science suggests that many types of programs and interventions have heterogenous effects (Duncan & Vandell, 2011; Weiss, Bloom, & Brock, 2013). We might expect several different patterns for the effects of curricula interventions, which may result from a number of characteristics including baseline characteristics of children, local area contexts, and program experiences. For example, past studies indicate that children at lower risk in terms of individual skills or characteristics (e.g., behavior problems) benefit less from earlier intervention than higher risk children (Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Bierman et al., 2010; Conduct Problems Prevention Research Group, 2011).

The majority of studies in developmental science examine treatment effect heterogeneity by examining subgroups defined by their baseline characteristics. For example, Miller, Farkas, Vandell, and Duncan (2014) found variation in the effects of Head Start in math and literacy

based on the levels of academic stimulation (low, average, high) in the home, which was measured before children entered the program. In a similar method of analysis, Watamura, Phillips, Morrissey, McCartney, and Bub (2011) examined children experiencing different levels of the quality of the home environment on childcare quality. They reported that children from low-quality home environments differentially benefited from high-quality child care. Evidence from the Sure Start program in the United Kingdom found that children from more advantaged backgrounds experienced greater benefits from the program (Belsky et al., 2006). Although these baseline subgroup analyses are useful, a different way to think about treatment effect heterogeneity is to consider children's level of achievement *after* the intervention if we think that the processes of the intervention might have differential effects for low- and high-achieving children.

A Distributional Perspective

When designing curricular interventions, we should consider not just the average impact of curricula on different subgroups of children, but also the way it affects the entire distribution of children's achievement as it could have important policy implications for understanding who benefits most from the curricula (Schochet, Puma, & Deke, 2014). In this case, we are interested in the degree to which the effects vary across the ex-post measures of achievement (e.g., outcome measures taken at the end of the intervention). To examine the effect of skill-specific curricula on the distribution of children's achievement in the current study, I take advantage of randomized assignment to treatment (skill-specific curricula) and control (business-as-usual curricula) classrooms to estimate quantile treatment effects (QTE; Firpo, 2007).

QTE allow for unconditional comparisons of the achievement distributions of children exposed to treatment and control curriculum classrooms and provide more information on the

nature of treatment effects on the treated sample than mean differences. In the context of experimental data, QTE are estimated by calculating the difference in the two marginal distributions (cumulative distribution functions, or CDFs) and are identified at each quantile akin to average treatment effects (Bitler, Hoynes, & Domina, 2014).[2] Recent distributional studies using this method have shown that many education programs and interventions do not have a uniform effect on all program recipients (Bitler et al., 2014; Penner, 2016). For example, findings from a reanalysis of the Head Start Impact Study suggest that Head Start attendance produced the largest gains in academic achievement for children at the bottom of the distribution (Bitler et al., 2014).

Estimating variation in treatment effects across the distribution of a posttest outcome measure may appear similar to examining variation in average treatment effects for subgroups of children based on their pretest scores. Experts on QTE (Bitler, Domina, Penner, & Hoynes, 2015; Schochet et al., 2014) offer some reasons why these sets of analyses might produce different results. Bitler et al. (2015) explain that estimating the treatment effects defined by the baseline score is "by level of achievement ex ante, which need not be the same as ex-post achievement" (p. 422). Additionally, Schochet et al. (2014) offer two reasons why results may differ. One reason is if a program participant's location in the distribution is changed by the treatment (e.g., rank preservation does not hold; Heckman, Smith, & Clements, 1997). For example, if an intervention has large positive effects for children with the lowest baseline scores and large negative effects for children with high baseline scores, then the distribution of the outcome variable could be reversed. In this scenario, the QTEs will show small impacts on any

[2]I use the term "quantile" as a flexible term for any cut of the distribution, however, to be more specific this study cuts the distribution at every percentile (from the 1st to the 99th). Both terms are used interchangeably throughout the chapter.

quantile of the outcome distribution, but the baseline subgroup analyses will show positive impacts for low-scoring children and negative impacts for high-scoring children. Another reason these estimates might differ is if there were changes in children's achievement that occurred after random assignment (i.e., after the pretest was administered) that could affect those with high and low pretest scores. For example, the quality of the classroom instruction or the implementation of the intervention could vary across the sample. This could potentially lead to a reordering of the posttest scores that would be captured in the QTE analysis but not in the subgroup analysis.

The current study provides a new way for thinking about the effects of curricula on the distribution of children's outcomes by reevaluating the same experimental data from the original PCER study to test whether the original null mean effects found are concealing important distributional differences. Given the large gaps in achievement between children of varying characteristics at school entry, paired with the widespread adoption of whole-child curricula that has shown to be relatively ineffective (Nguyen, 2017), it is of considerable policy interest to determine how skill-specific influence children's skills along the distribution. Understanding variation in curricula effects is critical for facilitating the most efficient use of limited resources, by informing decisions about how best to target specific curricular programs and suggesting ways to improve the design or implementation of the programs for high-quality early education settings.

Theorizing Heterogeneous Effects of Early Childhood Interventions

Several hypotheses might predict which types of students are most likely to benefit from any type of intervention that are informative to consider in the context of the current study. Outlined below are five different hypotheses of heterogeneous treatment effects. It is important to note that these hypotheses are not mutually exclusive in that there may be evidence of support

for multiple hypotheses at different quantiles of the distribution for key outcomes. Figure 4.1 provides an idealized illustration of the five hypotheses that might be expected when looking across the distribution of children's achievement.

[Insert Figure 4.1]

Hypothesis 1: Equal Effects. The equal effects hypothesis suggests that the intervention will influence children equally along the skills distribution. That is, the end-of-treatment effect of the curriculum intervention will be the same for children at every percentile.

Hypothesis 2: Compensatory Hypothesis. The compensatory hypothesis (Rutter, 1987; Sameroff & Chandler, 1975) predicts that skill-specific curricula will boost achievement for children primarily at the bottom of the skills distribution. These children derive the largest benefits from being in the intensive skill-building environments that may come with curricula targeting a specific content domain.

Hypothesis 3: Skills Beget Skills. The skills beget skills hypothesis (Cunha & Heckman, 2007) predicts that because academic skills are cumulative, skill-specific curricula will boost achievement for children primarily at the top of the skills distribution. This hypothesis mirrors the Matthew effect in developmental psychology (Stanovich, 1986). The skills children possess prior to the intervention allow them to gain further skills from the intervention. Children with higher initial skills will have a developmental advantage over children with lower skills, and as a result the gap between high- and low-achievers widens with the interventions.

Hypothesis 4: Goldilocks Hypothesis. This hypothesis suggests that curriculum interventions will be most effective for children at the middle of the distribution (i.e., average-achieving). Children with lower skills may not benefit as much from the intervention because there may be too much to learn that they cannot catch up in the classroom, and children with

higher skills may not see any benefit from the additional instruction provided by the curriculum thus experiencing a ceiling effect. Children at the middle of the skills distribution presumably receive just the right amount of instruction that they need from the intervention to experience the largest gains from the treatment. The Goldilocks hypothesis is supported in a study by Miller et al. (2014) who found that children receiving moderate support of parental academic stimulation experienced the largest gains in literacy by attending Head Start.

Hypothesis 5: No Effects. The no effects hypothesis suggest that skill-specific curricula have no effect across the distribution of achievement. Although the prior possibilities described above are all theoretically plausible, it may be the case that skill-specific curricula do not influence the distribution of achievement based on the results of earlier analyses of the PCER data (PCER, 2008), or that the skill-specific curriculum interventions did little to change children's learning environments.

Current Study

In the current study, I explicitly test these five hypotheses in the context of skill-specific preschool curriculum interventions. Specifically, I investigate the possibility that the weak average effects of the PCER study disguise larger program effects for high- or low-achieving students. I examine the variability in children's outcomes that were targeted by the skill-specific preschool curricula that were implemented. This study uses experimental data from PCER to examine the effect of literacy curricula on the distribution of children's achievement in preschool. The experimental nature of the PCER study also allow for causal estimates of this relationship. Prior research has only examined the average impacts of the experimental curricula (see Jenkins et al., 2018)—my study will use distributional methods to identify variation in effects that may have been previously hidden. I use quantile treatment effect estimation to test

the five hypotheses outlined above regarding the effects of skill-specific curricula on children's end-of-preschool achievement. As a complement to the QTE analyses, I also conduct baseline subgroup analyses by examining heterogeneous treatment effects separately for subgroups defined by the pretest scores.

The current study pools data from all grantees that implemented: 1) a literacy curriculum where the comparison condition was Creative Curriculum or High Scope—the two most commonly used curricular packages in early childhood classrooms; or 2) a literacy curriculum where the comparison condition was a locally-developed curriculum. These curricula are described in greater detail in the measures section and in the PCER report (2008). Table 4.1 provides a description of the different curriculum comparisons that will be made in this study as well as the reported impacts from the original PCER study.

[Insert Table 4.1]

Methods

Data

Beginning in 2003, 12 grantees around the United States were funded to study the effect of preschool curricula on children's academic and social-emotional outcomes into kindergarten in the PCER study. Curricula type in the PCER dataset include literacy, whole-child, or locally developed. Each grantee chose the study curricula for a total of 14 different curricula tested in 18 different locations. Research evaluation firms assisted with the evaluation of the curricula to ensure consistent data collection at each site, but each grantee was in charge of its own evaluation. Individual grantees were responsible for recruiting preschool centers to participate in the study. At each grantee site, either classrooms within preschool centers or entire centers themselves were randomly assigned to a treatment (experimental curriculum) or control

condition. For feasibility and to preclude cross contamination across classrooms, most research sites assigned only one curriculum to each center. Baseline data on children, parents, and preschools were collected in the fall of 2003, with post-treatment data collected in the spring of 2004. Approximately 2,900 children in 320 preschool classrooms participated in the study. The data include children who were either in Head Start, private child care, or public preschool.

The analyses in the PCER final report (2008) provide grantee-specific estimates of the standardized outcome differences between designated and counterfactual curricula. The subsample of the PCER data most relevant to the proposed study is the grantee sites and classrooms that used skill-specific literacy curricula. One PCER grantee included a math-focused preschool curriculum as the randomly assigned treatment compared with Creative Curriculum and High Scope, but this grantee was omitted due to power limitations. The inclusion criteria described above required me to drop six grantees and approximately 1,150 children from the study. The majority of children in the analysis sample in PCER were from low-income households and had mothers with less than a high school degree. Approximately half of the sample were female and the average age was approximately 4½ years old. About half of the sample were from families with at least one parent employed. Table 4.2 presents the descriptive characteristics of the participants in this study. It can be seen that the majority of the analytic sample scores below the nationally normed mean on the three of measures of language and literacy (discussed below).

[Insert Table 4.2]

Measures

Treatment Curricula. All literacy curricula focus on a literacy domain, which could include phonological skills, prewriting skills, or any other literacy skill. The instructional

strategies across the literacy curricula differed greatly. The treatment curricula in the first comparison include Early Literacy and Learning Model, Literacy Express, DLM Early Childhood Express supplemented with Open Court Reading Pre-K, Ready Set Leap, and Language-Focused Curriculum. Treatment curricula in the second comparison include Doors to Discovery, Let's Begin with the Letter People, and Bright Beginnings. Some of the curricular packages provided suggested literacy activities and materials; others had scripted curricula, included technology components, and provide teachers with activities for the entire day.

Comparison Curricula. Comparison curricula in this study include either Creative Curriculum and High Scope (comparison 1 sample) or locally-developed curricula (comparison 2 sample). Creative Curriculum and High Scope are considered "whole-child" curricula in that they both provides enriching experiences across the multiple domains of children's development (e.g., social-emotional, health, academic). Whole-child curricula typically take a constructivist approach to learning and emphasize child-centered active learning cultivated through strategic arrangement of the classroom environment (DeVries & Kohlberg, 1987; Piaget, 1976; Weikart & Schweinhart, 1987) and encourage children to interact independently with equipment, materials, and other children in the classroom. Locally-developed curricula consist of lesson plans and other materials developed less formally by preschool teachers and center directors. These curricula may sometimes include components of various commercial curricula.

Literacy and Language Outcomes. This study draws upon three literacy and language measures. The literacy measures, Letter-Word and Spelling, come from the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001). Children were assessed at the beginning and end of the preschool year. The Letter-Word subtest is similar to the PPVT in that it asks children to identify the letter or word spoken to them, and the test gradually

increases in difficulty to require the child to read words out of context. The Spelling subtest requires children to write and spell words presented to them. Both assessments from the WJ-III were administered by trained researchers and each took approximately 10 minutes to administer. Scores are standardized by the age of the child and nationally normed to have a mean of 100 and a standard deviation of 15.

The third measure is the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997), which assesses children's vocabulary knowledge. It takes approximately five to ten minutes to complete and is administered by a trained researcher. The measure requires the child to point to the picture that represents the word spoken to them by the researchers. Words increase in difficulty and scores are standardized according to the child's age. Like the Woodcock-Johnson subtests, the measure is nationally normed, with a mean of 100 and a standard deviation of 15. The analyses for this study will use standard scores from all three measures.

Child, Family, and Teacher Characteristics. A host of baseline characteristics of the child, family, and teacher are included in the analyses. At baseline, the maternal/primary caregiver reported on child and family demographic and background characteristics. Child-level characteristics include gender (1 = female), race (White as the reference group), age in months, and whether a non-English home language is spoken. Children's fall of preschool literacy and mathematics scores are also included in the analyses.[3] Maternal/primary caregiver and family characteristics include level of education, whether employed (1 = yes), age in years, log of

[3]This is not necessary to include with experimental data (see Krueger & Zhu, 2004), but I wanted to control for these observable baseline skills to adjust for any imbalances between the treatment and control groups. Bitler and colleagues (Bitler et al. 2015; 2016) follow the same method in generating their inverse propensity score weights. I ran models that excluded the baseline test scores from the inverse propensity score weights and found that it makes little difference for the final set of QTE analyses.

annual household income in thousands of dollars, and whether receiving welfare (1 = yes). Teacher characteristics included level of education and years of experience.

Analytic Plan

Before estimating the QTEs, I first need to address the lack of balance on baseline scores. I first use an inverse propensity score weighting approach as a nonparametric first step (Firpo, 2007), which allows me to balance baseline test scores across the two groups and to account for differences in the likelihood of being assigned to a classroom with literacy curricula or business-as-usual curricula. As an example, Figure 4.2 shows the CDFs and QTE plot for unweighted baseline WJ Spelling scores for literacy curricula classrooms compared with High Scope and Creative Curriculum classrooms. The CDFs present the spelling standard scores on the x-axis with the cumulative percent of the sample on the y-axis.

[Insert Figure 4.2]

The horizontal distance between these CDFs at each point in the distribution equals the difference in standard scores. This is the quantile treatment effect at that percentile. The corresponding QTE plot is also shown, where the x-axis represents the cumulative percentiles of the distribution and the y-axis represents the difference in standard scores between treatment and control classrooms at each percentile. The score difference (solid line) is plotted along with pointwise 90 percent confidence intervals (dashed lines), which are calculated by stratifying on the site location and treatment status and bootstrapping the estimates 999 times. These plots allow me to assess the degree to which randomization successfully balanced fall scores across the distribution of achievement on WJ Spelling. The QTE plot shows that children assigned to the treatment group were favored on the outcome. Between the 5th and 38th and the 63rd and 98th percentiles, there is a positive and significant difference between treatment and control where the

confidence intervals do not include zero, suggesting imbalance across the distribution in random assignment.

To estimate the inverse propensity score weights, I use a logistic regression model to predict assignment to a treatment or control curricula classroom as a function of baseline test scores for literacy and language and characteristics of the child, family, and teacher from Table 4.2. I calculate the predicted probability of being in the treatment group, $\hat{p}$, and construct weights of $1/\hat{p}$ for those in the treatment group and $1/(1-\hat{p})$ for the control group. As shown by the p-values for mean comparisons in Table 4.2, these weights balance the treatment and control groups on observable dimensions. I also ran a joint test of significance for these characteristics and found that it was not significant, suggesting that when using the inverse propensity score weights, there are no differences across random assignment groups.

Propensity score weighting allows me to obtain unconditional estimates while still adjusting for any post-randomization imbalance in baseline test scores. This is helpful in examining distributional differences, as it provides a semi-parametric way to adjust for many covariates while still allowing for unconditional comparisons (for a discussion, see Killewald & Bearak, 2014). That is, it allows me to account for differences in observable characteristics without examining conditional distributions. As shown in Figure 4.3, balance was achieved using these weights because the adjusted CDF plots between the treatment and control groups show that the distributions are now similar and the corresponding adjusted QTE plot show non-significance at every quantile. Thus, I will apply the inverse propensity score weights to estimate the unconditional spring of preschool QTEs for measures of children's literacy and language achievement.

[Insert Figure 4.3]

Using these CDFs, I examine the difference between these two distributions at various percentiles of the outcome variables. For example, I estimate the QTE at the .50 quantile, or 50th percentile, by subtracting the median test score of children in the control classrooms from the median test score of children in the treatment classrooms. By comparing test scores at a number of quantiles, I am able to observe the effect of skill-specific curricula on different portions of the distribution. If skill-specific curricula have different effects on relatively high-, average-, or low-achieving children, this method will identify these differences while mean comparisons with ordinary least squares (OLS) regression would not.

Results

Similar to the previous figures, Figures 4.4 through 4.6 show the test score differences between children in treatment and control classrooms (y-axis) for each percentile of the distribution (x-axis). When the solid blue line, which represents the point estimate at a given quantile, is above zero children in treatment classrooms are outscoring children in control classrooms, and when the solid line is below zero children in treatment classrooms are scoring lower than children in control classrooms. These differences are statistically significant when the area between the two dashed red lines, representing 90 percent confidence intervals, does not include the solid horizontal black line marking zero on the y-axis.

The quantile treatment effects of the two comparisons of treatment and control curricula reported below are organized by children's key outcomes targeted by the skill-specific curricula. To orient the reader to how I discuss the results for each outcome, I first report the average treatment effect (not shown on the QTE plot). Then, I discuss whether being in a literacy curriculum classroom improves children's skills on the outcome of interest by examining the point estimates (regardless of statistical significance) across the distribution on the QTE plot.

Finally, I discuss whether the gains associated with treatment are at the lower, middle, and/or top part of the test score distribution by highlighting parts of the distribution that are statistically significant.

WJ Letter-Word

The QTE results for the Letter-Word subtest in the spring of preschool are presented in Figure 4.4. For children in literacy classrooms compared with children in High Scope or Creative Curriculum classrooms (comparison 1), the average treatment effect is .15 SD. For the QTE analyses, the point estimates are positive across nearly the entire distribution of Letter-Word, except at the upper and lower tails. Between the 8th and 66th percentiles, the confidence intervals show these differences are statistically significant at the 10 percent level. Being in a classroom implementing a literacy curriculum leads to an increase in achievement on Letter-Word, with large and significant effects at the bottom of the distribution between the 7th and 22nd percentiles and somewhat smaller and significant effects between the 23rd and 51st percentiles. There are smaller and significant effects at the 52nd percentile, losing statistical significance beginning around the 67th percentile. The evidence is largely in favor of a compensatory effect of being assigned to a literacy curriculum classroom for those at the bottom and middle of the Letter-Word test score distribution.

For children in literacy classrooms compared with children in locally-developed curricula classrooms (comparison 2), average treatment effect is .10 SD. The QTE point estimates are positive between the 1st and 75th percentiles, in favor of children in the treatment classrooms. Letter-Word scores become negative or zero beginning at the top of the distribution, between the 76th and 99th percentiles, in favor of children in classrooms implementing locally-developed curricula. For this comparison, there is evidence of a compensatory effect for children at the

bottom of the Letter-Word distribution as the point estimates are larger and statistically significant between the 9th and 33rd percentiles.

[Insert Figure 4.4]

WJ Spelling

Figure 4.5 shows the average treatment effect and the QTE plot for spring Spelling subtest scores. For children in literacy classrooms compared with children in High Scope or Creative Curriculum classrooms, the average treatment effect is .20 SD. For the QTE plot, the point estimates of the effect of being in a treatment classroom are positive, with differences in standard score between treatment and control ranging from about 1 to 8 points. The point estimates are larger at the upper tail of the distribution, beginning around the 93rd percentile. The point estimates are statistically significant at some portions of the bottom of the Spelling test score distribution between the 2nd and 4th percentiles, 6th and 11th percentiles, and 23rd and 28th percentiles. Although a compensatory effect for children at the bottom of the distribution for Spelling cannot be ruled out, the QTE estimates at this portion of the distribution are inconsistent in terms of statistical significance. There is, however, consistent evidence of the skills beget skills hypothesis at the top of the Spelling test score distribution with the point estimates statistically significant beginning at the 82nd percentile, with differences between treatment and control classrooms as large as 8 standard score points.

For children in classrooms implementing a literacy curriculum compared with a locally-developed curriculum, the average treatment effect is .18 SD. The QTE plot shows that there are positive point estimates at the bottom tail that become negative between the 4th and 9th percentiles. Beginning at the 10th percentile, the point estimate differences remain consistently positive. Larger and statistically significant differences can be seen between the 71st and 94th

percentiles, with differences as large as 6 standard score points, consistent with the skills beget skills hypothesis.

[Insert Figure 4.5]

PPVT

The average treatment effect and the QTE results for the PPVT in the spring of preschool are shown in Figure 4.6. For children in literacy classrooms compared with children in High Scope or Creative Curriculum classrooms, the average treatment effect is -.01 SD. Point estimates are large and positive between the 1st and 12th percentiles—as large as 10 standard score points at the bottom tail. Beginning at the 14th percentile up until the 77th percentile, point estimates are negative and the largest differences are seen between the 47th and 54th percentiles where children in the literacy curricula classrooms are scoring lower on the PPVT than children in the locally-developed classrooms by about 3 standard score points. After the 79th percentile, point estimates are positive and are at or near zero. None of the point estimates are statistically significant, and thus lends support to the no effects hypothesis tested in this study. These patterns of results yield an average effect that is near zero, as shown on the green dashed line, and statistically insignificant.

Similar results were also found for children in literacy classrooms compared with children in locally-developed curricula classrooms. The average treatment effect is .06 SD. For the QTE plot, point estimates are positive, though near zero, for the bottom half of the distribution, between the 1st and 54th percentiles. Point estimates after the 56th percentiles are less consistent, with some parts of the distribution being positive, and in some other places negative. However, none of these differences is statistically significant thus rendering support for the no effects hypothesis.

[Insert Figure 4.6]

Baseline Skills Subgroup Analyses

Because it might be argued that examining treatment effect heterogeneity by conducting subgroup analyses defined by the pretest score is similar to the QTE analyses, I ran a complementary set of models examining the moderation between treatment status and baseline skills. Children's baseline skills were defined by cut points 1 SD above (high baseline score) and 1 SD below the mean (low baseline score). Children whose baseline scores were within 1 SD above and below the mean were the reference group (average baseline score). To be analogous to the QTE analyses as much as possible, I applied inverse propensity score weights to adjust for sample-specific differences at baseline in order to isolate curricular impacts. These weights were then included in the OLS regression models with standard errors clustered at the classroom level and fixed effects at the random assignment level which produces random assignment variation in the treatment and control groups.

Table 4.3 displays the subgroup results for the two comparisons on each of the three outcomes. In addition to estimating the main effect of children in the high and low baseline skills group (compared with the average baseline skills group) and the literacy curricula treatment effect, the models also include interactions between the treatment and each dummy indicator for the baseline skills group variables and the full set of covariates listed in the table's footnote. Below I report the general pattern of results only for the interactions since the purpose of the complementary analyses were to distinguish these results from the QTE results.

[Insert Table 4.3]

Results for WJ Letter-Word in the first two columns show that the literacy curricula treatment effect is largest for children who had low baseline scores, which is consistent with the

compensatory effect observed in the QTE analyses. Turning to the interaction variables for WJ Spelling, there is no statistically significant treatment effect for children with high or low baseline scores. However, there is a marginally significant ($p < .10$) treatment effect for children with high baseline scores but only for the first comparison sample of literacy curricula compared with High Scope or Creative Curriculum. Finally, the baseline skills subgroup analyses did not yield any significant interactions for high- or low-achieving children, confirming the QTE results.

Discussion

This study focused on testing five different hypotheses of the distributional effects in early childhood interventions using multisite experimental data of skill-specific curricula compared with business as usual curricula. I found the differential effects of skill-specific literacy curricula to depend on the outcome of interest. Specifically, I observed a compensatory effect in that children at the bottom of the distribution of Letter-Word benefited most from the treatment. I also found a skills beget skills effect for children at the top of the distribution on Spelling. For the domain of receptive vocabulary, as measured by the PPVT, I found that results were consistent with the no effects hypothesis. Because the three key measures of language and literacy were examined separately as opposed to being aggregated into a single composite score, it is important to note that children along the various test score distributions of each measure do not have to be the same individuals, although they may be. I also replicated the average treatment effects presented by Jenkins et al. (2018) and these findings complement and extend their analyses by using both baseline subgroup moderation analyses and QTE estimation to paint a more comprehensive picture of the children who benefit most from skill-specific literacy curricula interventions.

In the moderation analyses, I found that the results on children's WJ Letter Word and PPVT were consistent with results from the QTE analysis. Though it may not always be necessary to compare distributions at multiple points, it is a useful exercise for thinking about which points are worthy of comparison. In the case of children's WJ Spelling scores, I found largely null effects in the baseline subgroup analyses but positive and significant effects at the top of the distribution in the QTE analysis. The weak average (and statistically insignificant) effects were disguising larger treatment effects for high-achieving children. The results for this particular outcome might have well been missed had it instead focused on estimation of subgroup effects. It may be the case that there was some reordering of the posttest scores as a result of the intervention that could not be captured with the subgroup analyses, as suggested by Schochet et al. (2014). It could be that something about the classroom environment changed as a result of the intervention that affected the high-achieving students. It is important to note that while the baseline skills analyses and the QTE analyses were meant to be complementary to one another, they are not identical methods of estimation. These sets of analyses remain fundamentally different from one another in that QTEs are based on the posttest distribution while subgroup differences are based on the pretest distribution.

Though processes that occur at the center of the distribution and average group differences are of interest in evaluating interventions, these predictions based on OLS regressions assume that those at the top and bottom of the distribution are subject to identical processes. This is not necessarily a bad assumption, but in the context of this study considering non-mean-based comparisons, the evidence suggests that the processes affecting the low-achieving students are not the same as those affecting the high-achieving students. Perhaps the low-achieving children saw the largest gains on their letter-word scores as a result of the

intervention because they came in with little knowledge of the alphabet and these were basic skills that children could quickly pick up. For the spelling scores, it is possible that high-achieving children, having already acquired letter-word knowledge, are better able to piece together the letters of the alphabet to spell words. In any case, particular attention to specific and individualized instructional practices that are targeted to the skills of the child within these domain-focused curricular interventions seems warranted.

The current study provides a different way to understand variation in effects by examining how the outcomes of interest were influenced by the treatment across the outcome distribution of children's achievement at the end of the intervention. Examining heterogeneity of treatment effects for different children is important for understanding the modest, and sometimes null, effects of early childhood interventions. Educational and developmental researchers most commonly study this by looking at how well different subgroups of children are performing on the outcome; this study complements these traditional analyses with distributional methods and considers non-mean-based comparisons. The goal was to familiarize researchers with distributional analytic techniques, but it will still be important and policy-relevant for researchers to examine mean effects and heterogeneous effects on observable subgroups of children. Future studies interested examining in treatment effect heterogeneity will need to first ask what the potential sources of variation are in treatment effects to decide what is the most appropriate analytic technique.

Skill-based curricula are gaining much attention in the early childhood field, and we are seeing this with the adoption of these curricula in universal preschool programs in Boston (Weiland & Yoshikawa, 2013) and New York City (Mattera. & Morris, 2017; Morris, Mattera, & Maier, 2016). It will be important to assess whether early childhood interventions are having

their desired effects of boosting children's school readiness skills while narrowing achievement gaps—a question of great interest to the early childhood field and one with important policy implications. Future work will explore the degree to which effects within subgroups (i.e., dual language learners) are similar to those for the whole distribution. Additionally, it will be important to take a similar distributional approach to understanding variation in treatment effects for other skill-specific curricula, such as those targeting children's math skills or social-emotional behaviors, and whole-child curricula. Answering these questions will be key for future early childhood curriculum-based interventions in thinking about what works best for whom and under what circumstances.

References

Belsky, J., Melhuish, E. C., Barnes, J., Leyland, A. H., Romaniuk, H., & the National Evaluation of Sure Start Research Team. (2006). Effects of Sure Start local programmes on children and families: Early findings from a quasi-experimental, cross sectional study. *British Medical Journal, 332*, 1476–1748. doi:10.1136/bmj.38853.451748.2F

Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M.T., Lochman, J. E., McMahon, R. J., & Pinderhughes, E. (2010). The effects of a multiyear universal social-emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology, 78,* 156–168. doi:10.1037/a0018607

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology, 20*, 821–843. doi:10.1017/s0954579408000394

Bitler, M., Domina, T., Penner, E., & Hoynes, H. (2015). Distributional analysis in educational evaluation: A case study from the New York City voucher program. *Journal of Research on Educational Effectiveness*, *8*(3), 419-450. doi: 10.1080/19345747.2014.921259

Bitler, M. P., Hoynes, H. W., & Domina, T. (2014). *Experimental evidence on distributional effects of Head Start* (No. w20434). National Bureau of Economic Research.

Chambers, B., Cheung, A. C., & Slavin, R. E. (2016). Literacy and language outcomes of comprehensive and developmental-constructivist approaches to early childhood

education: A systematic review. *Educational Research Review*, *18*, 88-111. doi: 10.1016/j.edurev.2016.03.003

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, 38*(2), 136-163. doi: 10.2307/30034954

Clements, D., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*(2), 443-494. doi:1.3102/0002831207312908

Conduct Problems Prevention Research Group. (2011). The effects of the Fast Track preventive intervention on the development of conduct disorder across childhood. *Child Development, 82*, 331–345. doi:10.1111/j.1467-8624.2010.01558.x

Cunha, F., & Heckman, J. J. (2007). The technology of skill formation. *American Economic Review, 97*(2), 31–47. doi: 10.1257/aer.97.2.31

DeVries, R., & Kohlberg, L. (1987). *Programs of early education: The constructivist view.* Longman: White Plains, NY.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387-1388. doi:1.1126/science.1151148

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. The Journal of Economic Perspectives, 27(2), 109-132. doi:1.1257/jep.27.2.109

Duncan, G. J., & Murnane, R. J. (2014). *Restoring Opportunity: The crisis of inequality and the challenge for American education*. Cambridge, MA, and New York, NY: Harvard Education Press and Russell Sage Foundation.

Duncan, G. J., & Vandell, D. L. (2011). Understanding variation in the impacts of human capital interventions on children and youth. Working Paper, Irvine Network on Interventions in Development.

Dunn, L. M., & Dunn, L. M. (1997). Peabody Picture Vocabulary Test-Third Edition (PPVT-III). Upper Saddle River, NJ: Pearson Publishing.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, *75*(1), 259-276. doi: 10.1111/j.1468-0262.2007.00738.x

Hamre, B. K., Downer, J. T., Kilday, C. R., & McGuire, P. (2008). *Effective teaching practices for early childhood mathematics*. White paper prepared for the National Research Council.

Jenkins, J.M., Duncan, G.J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. (2018). Boosting school readiness: Should preschools target skills or the whole child? *Economics of Education Review*. doi: 10.1016/j.econedurev.2018.05.001

Justice, L. M., McGinty, A. S., Piasta, S. B., Kaderavek, J. N., & Fan, X. (2010). Print-focused read-alouds in preschool classrooms: Intervention effectiveness and moderators of child outcomes. *Language, Speech, and Hearing Services in Schools*, *41*(4), 504-520. doi: 10.1044/0161-1461(2010/09-0056)

Killewald, A., & Bearak, J. (2014). Is the motherhood penalty larger for low-wage women? A comment on quantile regression. *American Sociological Review*, *79*(2), 350-357. doi: 10.1177/0003122414524574

Klein, L., & Knitzer, J. (2006). *Effective preschool curricula and teaching strategies*. New York, NY: National Center for Children in Poverty.

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing, 24*(3), 305-337. doi: 10.1007/s11145-009-9214-6

Miller, E. B., Farkas, G., Vandell, D. L., & Duncan, G. J. (2014). Do the effects of Head Start vary by parental preacademic stimulation? *Child Development*, *85*(4), 1385-1400.

Mattera, S. K., & Morris, P. (2017). *Making Pre-K Count and High 5s: Preliminary kindergarten impacts*. New York, NY: MDRC.

Morris, P. A., Mattera, S. K., & Maier, M. F. (2016). *Making Pre-K Count: Improving math instruction in New York City*. New York, NY: MDRC.

Nguyen, T. (2017). Impacts of targeted and global preschool curricula on children's school readiness: A meta-analytic review. Working Paper, Irvine Network on Interventions in Development.

Penner, E. K. (2016). Teaching for all? Teach For America's effects across the distribution of student achievement. *Journal of Research on Educational Effectiveness, 9*(3), 259-282. doi: org/10.1080/19345747.2016.1164779

Piaget, J. (1976). *Piaget's theory*. New York, NY: Springer.

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness*. (NCER 2008-2009). Washington, DC: National Center for Education Research, Institute for Education Sciences, U.S. Department of Education.

Sameroff, A. J., & Chandler, M. J. (1975). Reproductive risk and the continuum of caretaker casualty. In F. D. Horowitz (Ed.), *Review of child development research* (Vol. 4) (pp. 187–244). Chicago, IL: University of Chicago Press.

Schochet, P. Z., Puma, M., & Deke, J. (2014). *Understanding variation in treatment effects in education impact evaluations: An overview of quantitative methods* (NCEE 2014–4017). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development.

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

Watamura, S. E., Phillips, D. A., Morrissey, T. W., McCartney, K., & Bub, K. (2011). Double jeopardy: Poorer social-emotional outcomes for children in the NICHD SECCYD

experiencing home and child-care environments that confer risk. *Child Development, 82,* 48-65. doi:10.1111/j.1467-8624.2010.01540.x

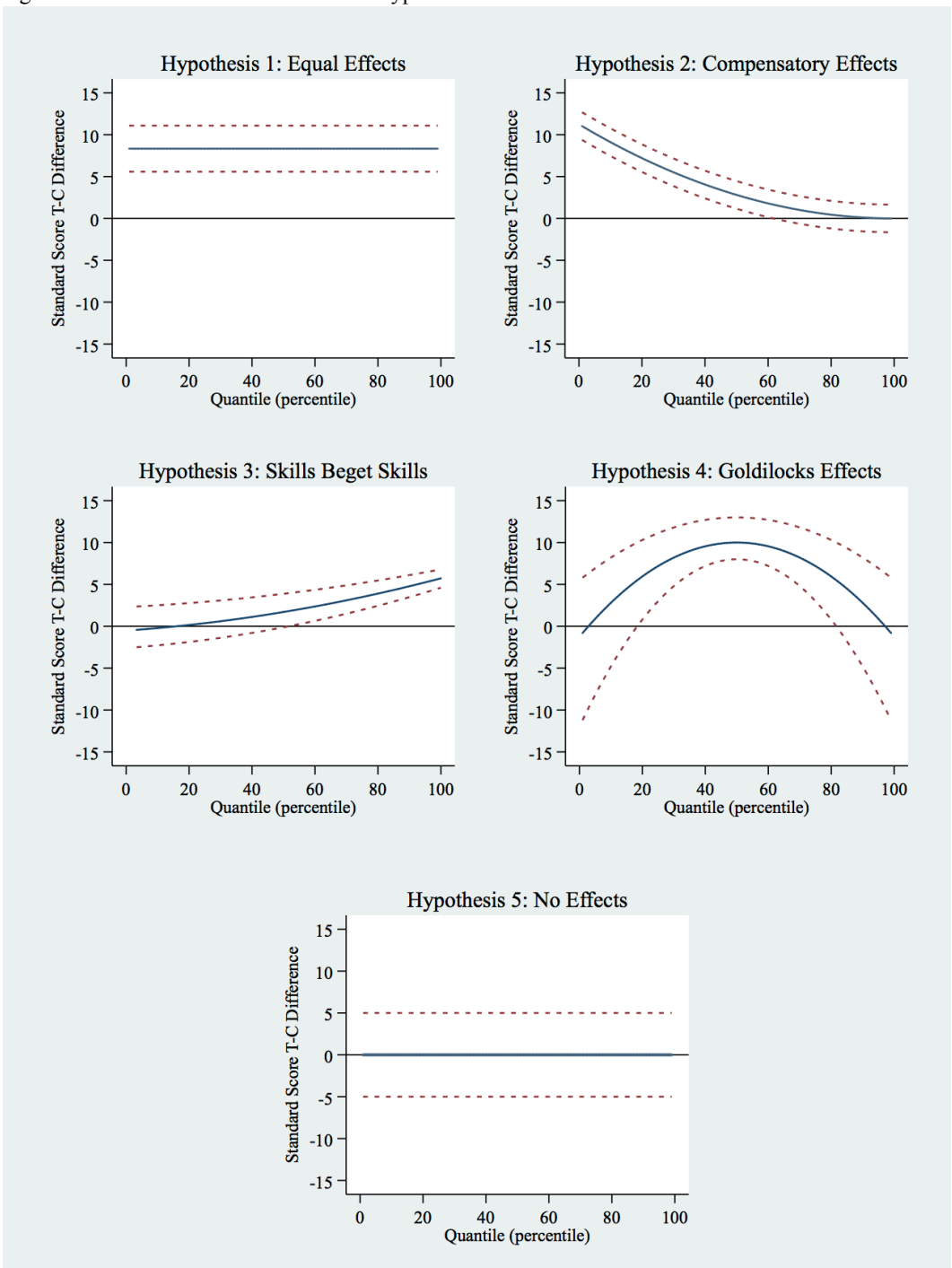
Weikart, D. P., & Schweinhart, L. (1987). The High/Scope cognitively oriented curriculum in early education. In J. L. Roopnarine & J. E. Johnson (Eds.), *Approaches to early childhood education* (pp. 253-268). New York: Merrill/Macmillan.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*(6), 2112-213. doi: 10.1111/cdev.12099

Weiss, M. J., Bloom, H. S., & Brock, T. (2013). *A conceptual framework for studying the sources of variation in program effects (MDRC Working Paper on Research Methodology).* New York, NY: MDRC.

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

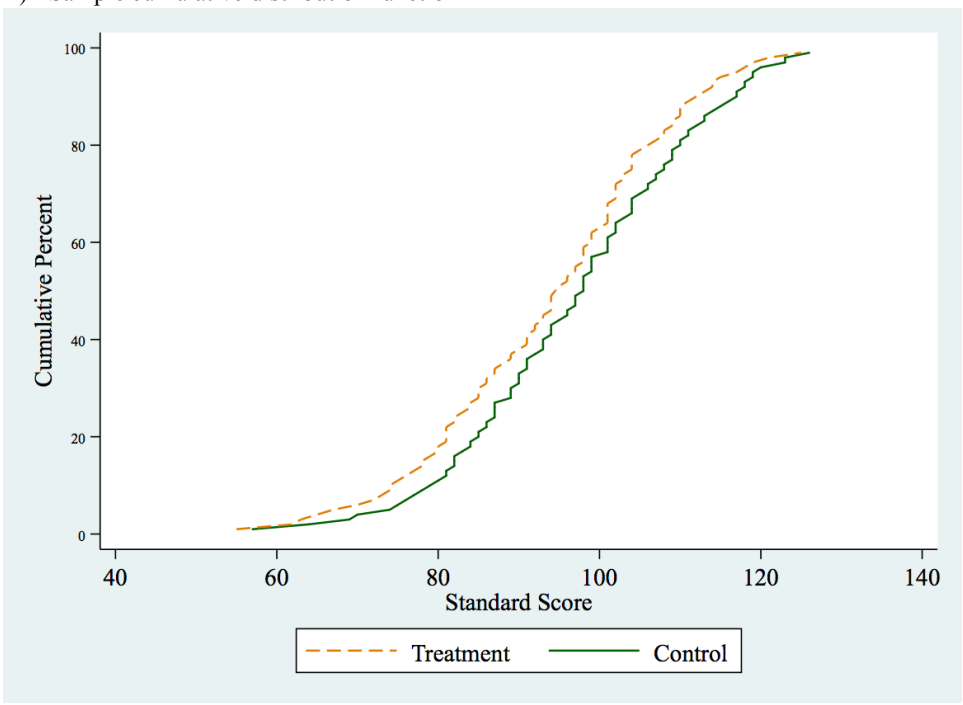
Figure 4.1. Idealized distributions of five hypotheses of effects of curriculum interventions.



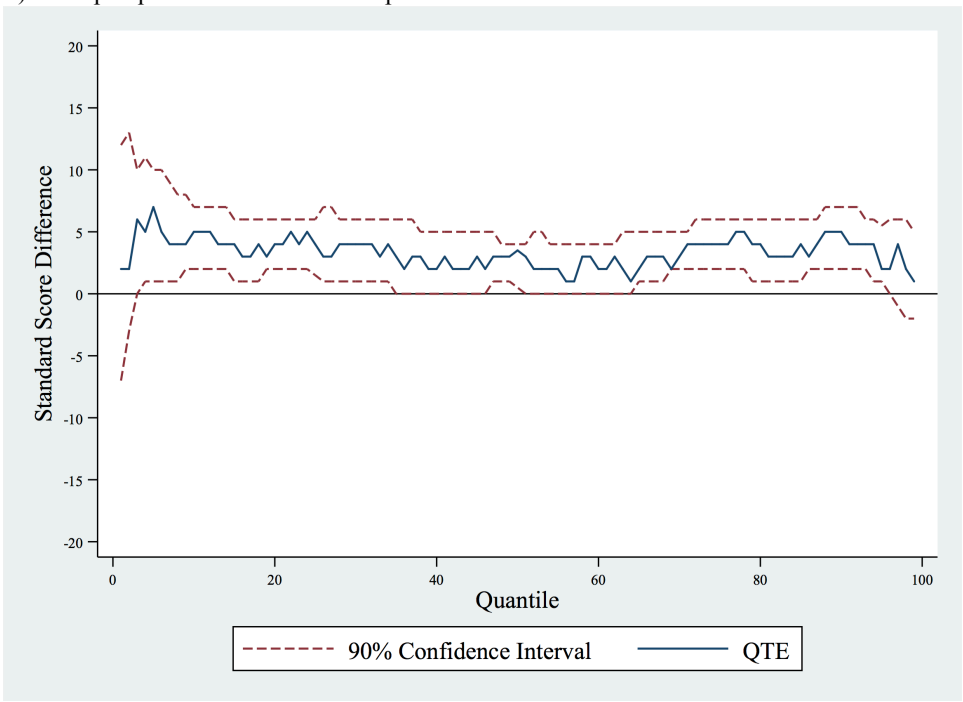
Note. Solid line shows idealized quantile treatment effect plots for five different hypotheses of distributional effects of curriculum interventions. Top and bottom dotted lines represent 90% confidence intervals.

Figure 4.2. Example of randomization imbalance in the PCER study.

A) Sample cumulative distribution function



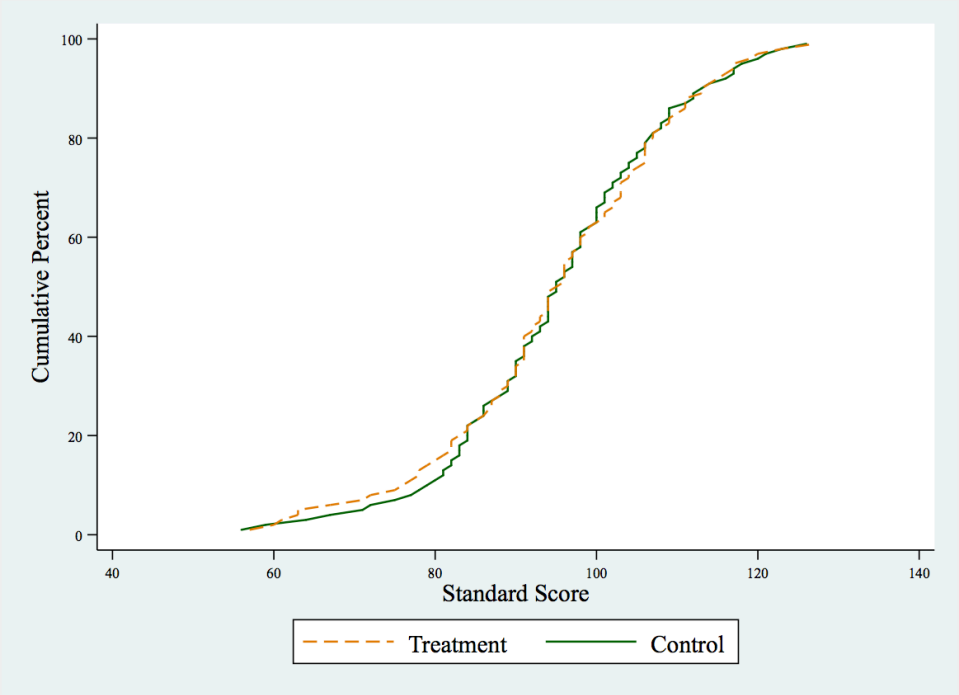
B) Sample quantile treatment effect plot



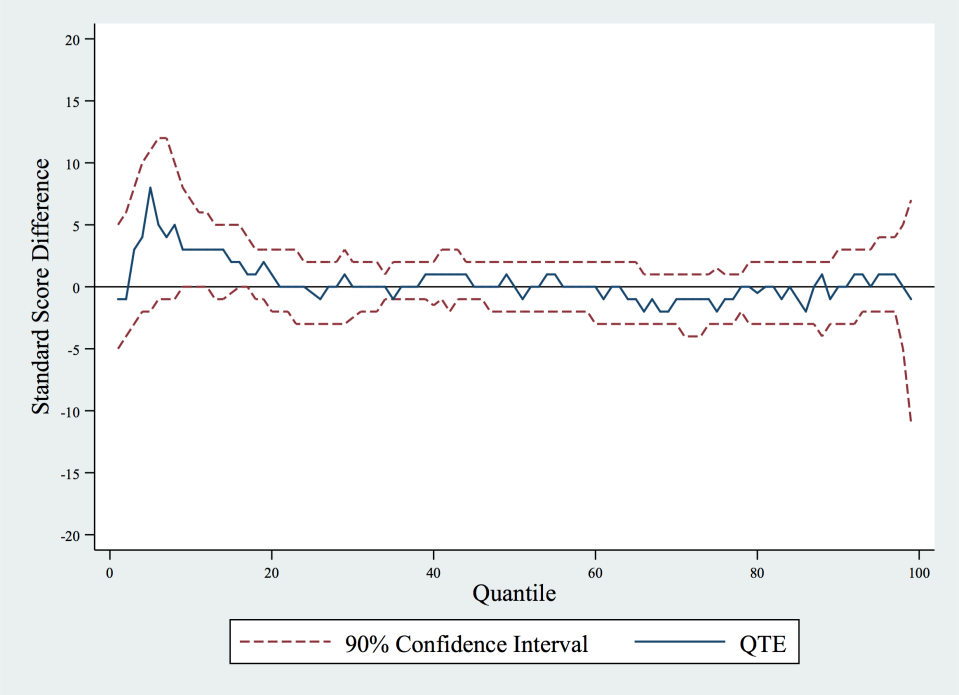
Note. This example is for comparison 1 of literacy curricula classrooms compared with High Scope or Creative Curriculum classrooms on Woodcock-Johnson Spelling subtest standard scores at baseline (fall). Estimates are weighted.

Figure 4.3. Example of balance across treatment and control achieved from inverse propensity score weighting of distributions.

A) Sample cumulative distribution function



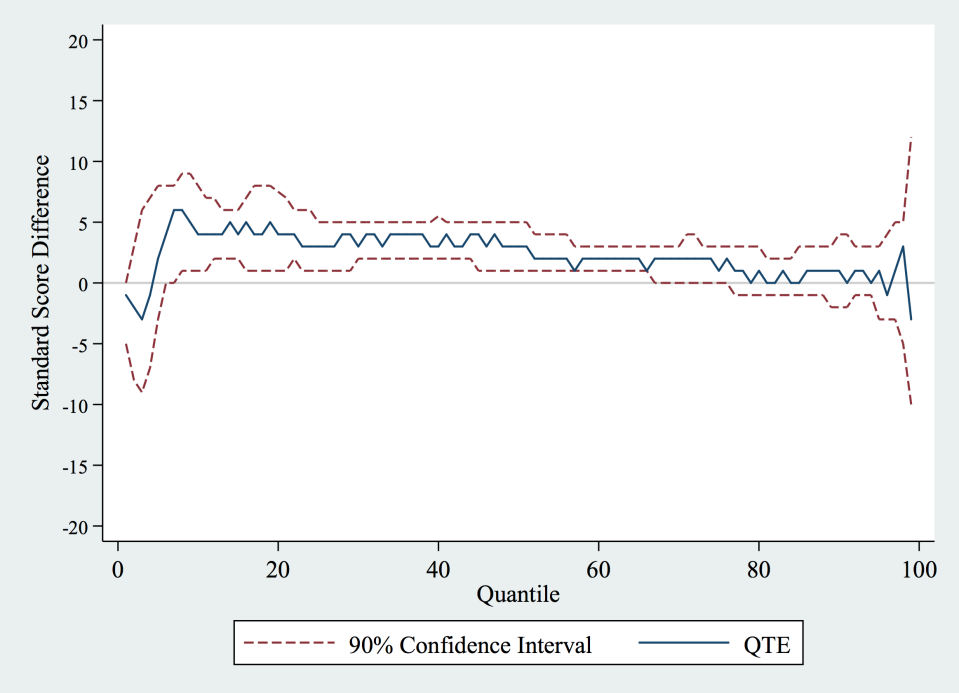
B) Sample quantile treatment effect plot



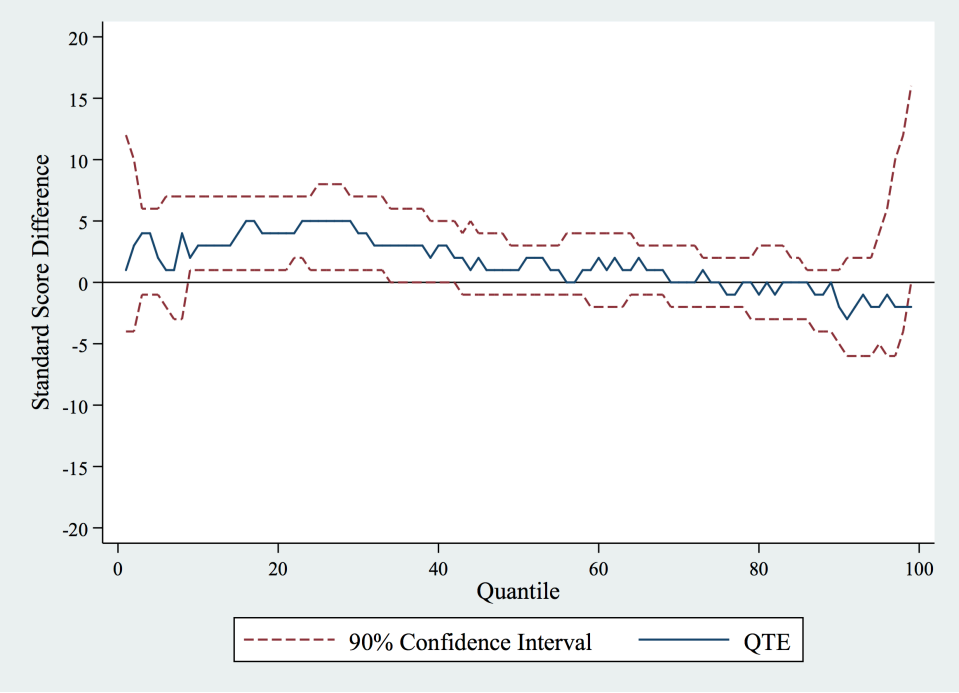
Note. This example is for comparison 1 of literacy curricula classrooms compared with High Scope or Creative Curriculum classrooms on Woodcock-Johnson Spelling subtest standard scores at baseline (fall).

Figure 4.4. Quantile treatment effect plots for Woodcock-Johnson Letter-Word.

A) Comparison 1: Literacy curricula compared with whole-child curricula



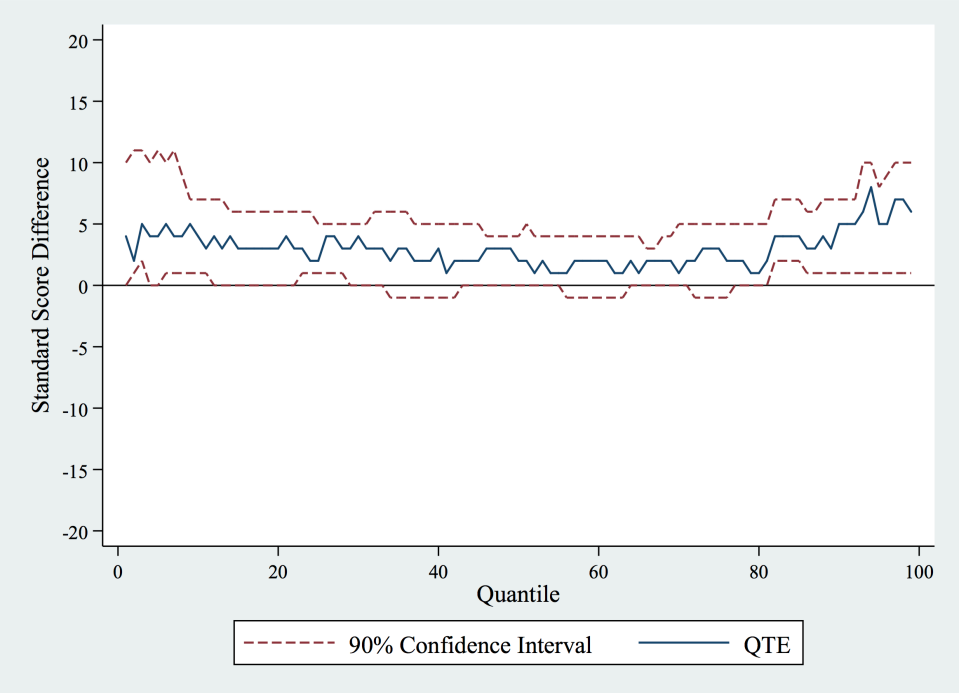
B) Comparison 2: Literacy curricula compared with locally-developed curricula



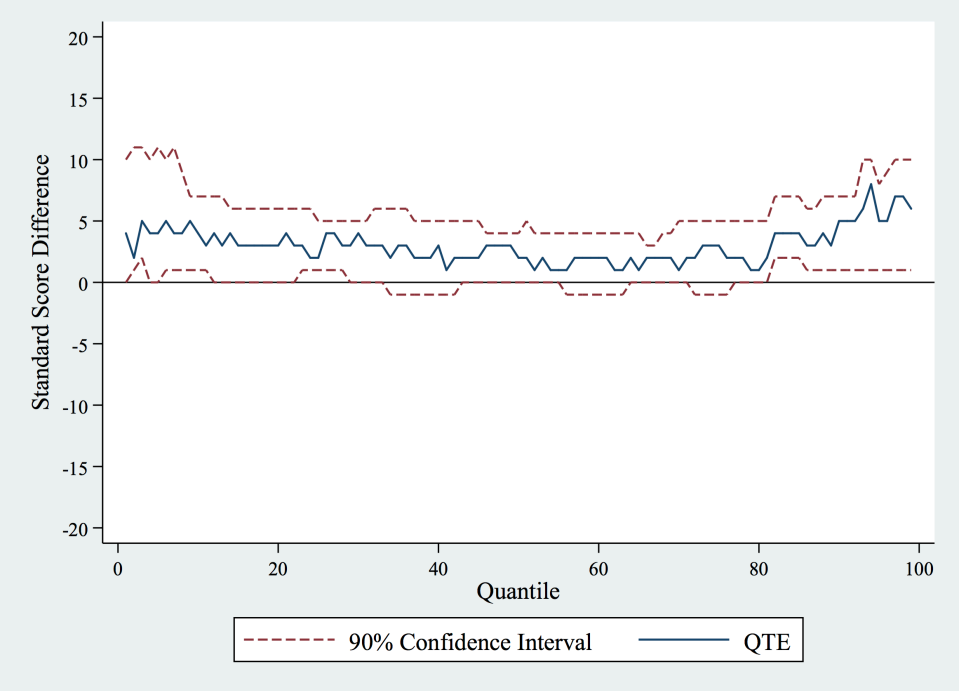
Note. Estimates are weighted using inverse propensity score weights. Weights are $1/\hat{p}$ for treatment observations and $1/(1-\hat{p})$ for control observations, where $\hat{p}$ is generated from a logistic regression of treatment status on baseline child, family, and teacher demographics, sample design variables, and baseline test scores on literacy and language outcomes. 90% confidence intervals are obtained by bootstrapping with replacement 999 times within randomization site.

Figure 4.5. Quantile treatment effect plots for Woodcock-Johnson Spelling.

A) Comparison 1: Literacy curricula compared with whole-child curricula



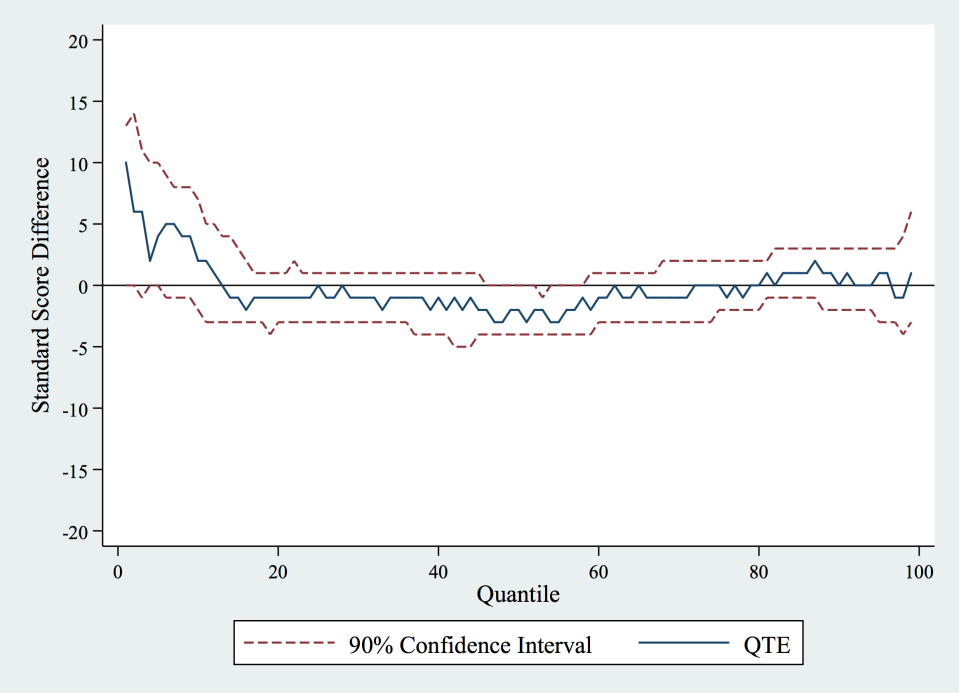
B) Comparison 2: Literacy curricula compared with locally-developed curricula



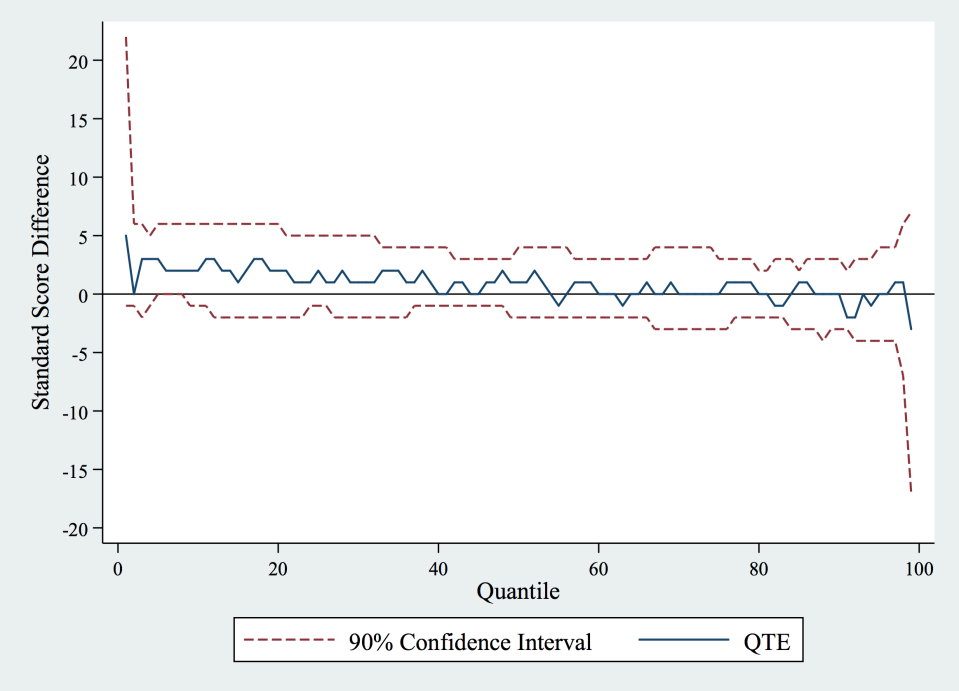
Note. Estimates are weighted using inverse propensity score weights. Weights are $1/\hat{p}$ for treatment observations and $1/(1-\hat{p})$ for control observations, where $\hat{p}$ is generated from a logistic regression of treatment status on baseline child, family, and teacher demographics, sample design variables, and baseline test scores on literacy and language outcomes. 90% confidence intervals are obtained by bootstrapping with replacement 999 times within randomization site.

Figure 4.6. Quantile treatment effect plots for Peabody Picture Vocabulary Test.

A) Comparison 1: Literacy curricula compared with whole-child curricula



B) Comparison 2: Literacy curricula compared with locally-developed curricula



Note. Estimates are weighted using inverse propensity score weights. Weights are $1/\hat{p}$ for treatment observations and $1/(1-\hat{p})$ for control observations, where $\hat{p}$ is generated from a logistic regression of treatment status on baseline child, family, and teacher demographics, sample design variables, and baseline test scores on literacy and language outcomes. 90% confidence intervals are obtained by bootstrapping with replacement 999 times within randomization site.

Table 4.1. Description of curricula comparisons in the analytic sample.

Grantee	Children	Site	Treatment Curriculum	Control Curriculum(a)	Project-reported Impacts	
					Literacy	Math
Comparison 1: Literacy curricula v. High Scope and Creative Curriculum						
University of North Florida	Treatment: 140 Control: 110	FL	Early Literacy and Learning Model	Creative Curriculum	ns, ns, ns	ns, ns
Florida State University	Treatment: 100 Control: 100	FL	Literacy Express	High Scope	ns, ns, ns	ns, ns
Florida State University	Treatment: 100 Control: 100	FL	DLM Early Childhood Express supplemented with Open Court Reading Pre-K	High Scope	+, +, +	+, ns
University of California, Berkeley	Treatment: 150 Control: 140	NJ	Ready, Set, Leap	High Scope	ns, ns, ns	ns, -
University of Virginia	Treatment: 100 Control: 100	VA	Language Focused	High Scope	ns, ns, ns	ns, ns
Comparison 2: Literacy curricula v. Locally-developed curricula						
University of Texas Health Science Center at Houston	Treatment: 100 Control: 100	TX	Doors to Discovery	Locally-developed	ns, ns, ns	ns, ns
University of Texas Health Science Center at Houston	Treatment: 100 Control: 100	TX	Let's Begin with the Letter People	Locally-developed	ns, ns, ns	ns, ns
Vanderbilt University	Treatment: 100 Control: 110	TN	Bright Beginnings	Locally-developed	ns, ns, ns	ns, ns

Note. This table replicates what is described in Jenkins et al. (2018). Observations in the second column are rounded to the nearest 10 in accordance with NCES restricted data requirements. Literacy was measured with Woodcock-Johnson Letter Word Identification, Woodcock-Johnson Spelling, and Peabody Picture Vocabulary Test, respectively. Math was measured with the Woodcock-Johnson Applied Problems and the Child Mathematics Assessment-Abbreviated, respectively. ns = not significant; + beneficial impact with $p<.05$; - detrimental impact with $p<.05$. All literacy curricula focused on a literacy domain, which could have included phonological skills, prewriting skills, or any other literacy skill.

Table 4.2. Sample characteristics of analytic sample.

	Comparison 1					Comparison 2				
	Observed Group Means		Weighted Group Means			Observed Group Means		Weighted Group Means		
	Literacy	High Scope or Creative	Literacy	High Scope or Creative	Diff	Literacy	Locally-developed	Literacy	Locally-developed	Diff
Child and family characteristics										
Female	0.46	0.45	0.45	0.47	0.02	0.45	0.49	0.42	0.45	0.03
White	0.24	0.31	0.28	0.31	0.03	0.52	0.32	0.40	0.40	0.00
Black	0.60	0.54	0.57	0.54	0.03	0.10	0.45	0.29	0.29	0.00
Hispanic	0.10	0.11	0.10	0.10	0.00	0.31	0.16	0.22	0.23	0.01
Asian/Native/Other	0.05	0.04	0.04	0.05	0.01	0.07	0.07	0.08	0.08	0.00
Child age	4.56	4.58	4.57	4.57	0.00	4.56	4.55	4.59	4.56	0.03
Mother has high school degree	0.54	0.61	0.57	0.58	0.01	0.43	0.51	0.45	0.47	0.02
Mother has some college	0.35	0.30	0.34	0.32	0.02	0.28	0.37	0.35	0.34	0.01
Mother has BA or higher	0.11	0.08	0.10	0.10	0.00	0.30	0.12	0.20	0.19	0.01
Mother working	0.70	0.66	0.68	0.69	0.01	0.60	0.59	0.60	0.59	0.01
Mother age	30.77	30.52	30.68	30.66	0.02	32.89	32.74	32.79	32.77	0.02
Mother married	0.38	0.34	0.37	0.36	0.01	0.72	0.55	0.60	0.62	0.02
Log of annual household income	3.01	2.93	2.97	2.99	0.02	3.42	2.98	3.18	3.17	0.01
Receiving welfare	0.14	0.19	0.14	0.17	0.03	0.08	0.13	0.09	0.11	0.02
Non-English home language	0.05	0.05	0.05	0.05	0.00	0.21	0.15	0.17	0.18	0.01
Teacher characteristics										
No college	0.27	0.14	0.23	0.20	0.03	0.10	0.23	0.15	0.16	0.01
Associates	0.08	0.09	0.08	0.08	0.00	0.06	0.16	0.09	0.13	0.04
Bachelors	0.50	0.67	0.56	0.58	0.02	0.61	0.38	0.56	0.47	0.09
Female	0.98	0.97	0.98	0.98	0.00	0.97	1.00	0.98	1.00	0.02
Age	41.39	40.68	41.00	40.96	0.04	39.38	38.60	38.42	38.49	0.07
Years teaching experience	12.32	11.58	11.79	11.76	0.03	14.41	11.76	13.70	12.61	0.09
Child outcomes at pre-k entry										
PPVT standard score	87.80	87.44	87.68	87.60	0.08	90.49	85.98	88.02	87.97	0.05
WJ Letter-Word standard score	101.27	98.67	100.27	100.34	0.07	98.96	97.64	98.22	98.17	0.05
WJ Spelling standard score	95.04	94.47	94.88	94.84	0.04	92.29	92.32	92.27	92.31	0.04

Note. Diff refers to differences between the observed and the propensity score weighted proportions or means, where $p < .10$.

Table 4.3. Key regression coefficients for models interacting baseline skills group with treatment condition.

	WJ Letter-Word		WJ Spelling		PPVT	
	Comparison 1: Literacy curricula v. High Scope and Creative Curriculum	Comparison 2: Literacy curricula v. Locally developed	Comparison 1: Literacy curricula v. High Scope and Creative Curriculum	Comparison 2: Literacy curricula v. Locally developed	Comparison 1: Literacy curricula v. High Scope and Creative Curriculum	Comparison 2: Literacy curricula v. Locally developed
Treatment (assignment to classroom using literacy curricula)	0.15*	0.13+	0.17*	0.15+	-0.01	0.06
	(0.08)	(0.07)	(0.07)	(0.09)	(0.07)	(0.08)
Low baseline score (below 1 SD of the mean)	0.23*	0.16*	0.09	0.13	-0.11	0.11
	(0.09)	(0.07)	(0.11)	(0.09)	(0.07)	(0.09)
High baseline score (above 1 SD of the mean)	0.12	0.09	0.13+	0.11	0.12	-0.09
	(0.09)	(0.10)	(0.07)	(0.08)	(0.08)	(0.09)
Treatment x Low baseline score	0.20*	0.14*	0.16	0.12	-0.02	-0.07
	(0.09)	(0.08)	(0.11)	(0.08)	(0.07)	(0.07)
Treatment x High baseline score	0.12	0.13	0.15+	0.10	0.12	-0.04
	(0.10)	(0.10)	(0.09)	(0.08)	(0.09)	(0.10)
Observations	890	480	890	480	890	480

Note. Outcome variables are all standardized. Standard errors in parentheses are clustered at the classroom level. Fixed effects at the random assignment site level are included in all models. Inverse propensity score weights were applied to all models. Treatment was modeled using child, family, and teacher characteristics. Child and family background characteristics include: child's gender, race, and age (in months). Family background characteristics include: parent or primary caregiver's highest level of education (in years), annual household income (in thousands), whether receiving welfare. Teacher characteristics include: gender, race, level of education (with the reference group as no degree), salary (in thousands), age (in years), and teaching experience (in years). Reference group for the baseline score is the average group scoring within 1 SD below or above the mean. WJ = Woodcock-Johnson; PPVT = Peabody Picture Vocabulary Test. Number of child observations were rounded to the nearest 10 in accordance with NCES data policies. +p<.10; *p<.05; **p<.01.

CHAPTER FIVE

Summary and Conclusion

If a key purpose of preschool programs is to promote children's school readiness, it is essential that we know whether and how curricula contribute to improved classroom processes and children's development of across a host of concrete skills, including literacy, language, numeracy, self-regulation, and social-emotional behaviors. Frede and Ackerman (2007) write in a policy brief for the National Institute for Early Education Research (NIEER) that "the promise of preschool will not be met if the curricula are not rigorously designed, carefully researched, and implemented as intended" (p. 12). Within this dissertation, I aimed to address some of the issues surrounding curricula in early childhood education: 1) the differential relations of widely used "research-based" whole-child preschool curricula packages and classroom quality and children's school readiness skills, 2) how such a curriculum is implemented in a real-world classroom context, and 3) the heterogeneous treatment effects of skill-specific curricula, by using different sources of data and descriptive, qualitative, and quasi-experimental methods.

This dissertation was largely motivated by work I have done in Nguyen (2017). Figure 5.1 from my meta-analysis shows the relative impacts of various curricula on children's language, literacy, and math outcomes. Literacy curricula effect sizes ranged from .17 to .20 SD and math curricula effect sizes ranged from .39 to .46 SD for the targeted content domain, when compared with whole-child and locally-developed curricula, respectively. The effect size for whole-child curricula compared with locally-developed curricula was a statistically insignificant .08 SD on an academic composite of children's outcomes (language, literacy, and math). Chapters 2 and 3 focused specifically on this last effect size, illustrated on the last vertical bar plot of Figure 5.1, for whole-child curricula compared with locally-developed curricula. In

Chapter 2, I aimed to get a better understanding of the relative associations between various whole-child curricula, a subset of which were included in the calculation of that effect size, and classroom quality and children's outcomes. In Chapter 3, I conducted qualitative classroom observations with the goal of answering the question of "what a whole-child curriculum even looks like in practice?" In Chapter 4, I focus on the first two bar plots displaying the meta-analytic *average* effect sizes of literacy skill-specific curricula, ranging from .17 to .20 SD. I draw upon data from a subset of curriculum interventions that were included in the calculation of those effect sizes to estimate heterogeneous treatment effects.

[Insert Figure 5.1]

Below, I present results regarding the relative associations of various curricula, the extent to which they are implemented, and for whom they are most effective. Following the summary of findings, I discuss some of the themes and key lessons learned that emerged as well as the substantive, methodological, and policy implications of this work. Finally, I close with a discussion of new directions for future research, both in general and for my own trajectory, as informed by this dissertation.

Summary of Findings

In Chapter 2, I examined associations of various whole-child curricula on classroom quality and children's school readiness outcomes. Using data from the Head Start Family and Child Experiences Survey, 2009 Cohort, I find that none of the widely used whole-child curricula in Head Start classrooms improve classroom quality or children's school readiness outcomes. One curriculum, High Reach, actually scored lower on measures of classroom quality and children in those classrooms scored lower on both academic and social skills outcomes than children in Creative Curriculum classrooms. Overall, these results suggest that there are few

distinguishing characteristics of whole-child curricula most commonly used in Head Start programs today.

Next, in Chapter 3, I focused on evaluating the extent to which teachers implement a whole-child curriculum across different classrooms through open-ended observations of fidelity. Using a theory-driven evaluation framework, I evaluate how teachers tailor and implement the High Scope curriculum with fidelity and/or ways in which implementation was limited or absent when applied to different children and classroom settings. My findings suggest that teachers are in fact implementing components of High Scope, including plan-do-review, child-centered active learning, and promoting learning across all developmental domains. However, some elements of the curriculum that are not as explicit and straightforward such as scaffolding children's learning are more difficult for teachers and point to the need for more supports in implementing the curriculum's instructional practices faithfully.

In Chapter 4, I reanalyzed the PCER data to examine whether the original weak average effects were disguised by larger intervention effects for different children at various points along the outcome distribution. A rigorous econometric technique, quantile treatment effects estimation, was used to identify this variation in treatment impacts. I find that skill-specific literacy curricula have differential effects at various points across the distribution of children's literacy and language skills depending on the developmental outcome. For children's letter-word outcomes, those at the bottom of the distribution benefited most from the treatment, consistent with the *compensatory* hypothesis. For children's spelling outcomes, those at the top of the distribution benefited most from the treatment, consistent with the *skills beget skills*, or *accumulated advantages,* hypothesis. For children's receptive vocabulary, no heterogenous treatment effects were found across the distribution of achievement.

Key Themes and Lessons Learned

Three independent studies addressed the overarching questions of how various preschool curricula are more or less effective in promoting classroom quality and children's school readiness. Taken together, they have important substantive, methodological, and policy implications. Several key themes and lessons prevailed across the studies and each of these is described below.

What does it mean to be a research-based curriculum? The whole-child curricula examined in Chapters 2 and 3 are described as "research-based" by their publishers, but few actually explain these claims. Though the format and method of whole-child curricula may be supported by research, there is little to no research indicating that these curricula have any effect on the classroom environment or children's school readiness. Substantively, my dissertation shows that when we compare these various and widely adopted "research-based" whole-child curricula to each other, there is no evidence of differential associations with classroom quality and children's outcomes. One potential reason why these whole-child curricula might have shown null to adverse associations is that they do not have specific learning goals. Research-based curricula include well-designed, engaging learning experiences that support specific goals and objectives (Chambers et al., 2010). Clear and well-designed activities prompt and encourage teachers to use various strategies and interactive materials to engage children in their learning and development. Further, the depth of each learning domain needs to be made clear through an organized scope and sequence, or a curricular map that identifies how the curriculum covers each domain element in a sequenced, progressive manner. However, a closer look at High Scope and Creative Curriculum, the two most commonly used preschool packages included in Chapter 2, suggests that there is limited evidence of a defined scope or sequence.

Though it may be the case that these curricula on their own may be effective, but just not differentially effective as shown in Chapter 2, it is important to note that claims that a curriculum is based on research should be carefully scrutinized to reveal the nature and extent of the connection between the two (Clements, 2007). From a policy perspective, preschool program decision-makers need to investigate the research base for any curriculum given the substantial costs associated with purchasing a package—upwards of $2,000 per classroom. These decision-makers should also be wary of curriculum developers' claims unless they are confirmed by researchers who are unaffiliated with the curriculum (Frede & Ackerman, 2007). The policy debate regarding the effectiveness of preschool curricula would benefit from more rigorous, telling empirical research.

Understanding fidelity of implementation is key. Questions around fidelity of implementation plague curricula research, and this implementation aspect has often been overlooked in outcome-focused studies (Durlak, 2010; Durlak & DuPre, 2008; Gilliam, Ripple, Zigler, & Leiter, 2000; McCall, 2009). If we are going to rely on research to help preschool programs distinguish between the impact of a strong and weak curricula, we have to know that the research actually assesses the use of the curriculum in the classroom. To what degree teachers seem to be using the curriculum as intended has not been adequately captured in large-scale datasets such as FACES in Chapter 2 or PCER in Chapter 4. Implementing a curriculum can be challenging, and programs often must train and mentor teachers to implement the chosen curriculum faithfully; however, we cannot expect that each teacher will implement a curriculum identically across classrooms (Domitrovich & Greenberg, 2000; Dusenbury, Brannigan, Hansen, Walsh, & Falco, 2005; Justice, Mashburn, Hamre, & Pianta, 2008). This is especially true for a whole-child curriculum, which is meant to be flexible and tailored to children.

I undertook open-ended classroom observations in Chapter 3 to understand the dynamics associated with whole-child curriculum implementation in a real-world context. Preschool teachers in Head Start are mandated to use the prescribed curriculum as a framework to guide interactions and activities in the classroom but must also meet the demands of high-stakes policy contexts related to program planning, evaluation, and accountability. The observations revealed that teachers implement the explicit, outlined parts of the High Scope curriculum with fidelity, but struggle with scaffolding children's learning, which may be more abstract and require more skill on their part. Because teachers are "street level" implementers of children's daily experiences in the classroom (Lipsky, 1980), it is important to understand the particular challenges teachers may face in implementing a curriculum with fidelity. My observational findings are the beginning of an evidence base to identify what specific areas of implementation support teachers need in order to realize the potential of curricular impact on promoting children's school readiness.

Heterogeneity of treatment impacts may help explain null findings of curriculum-based interventions. Methodologically, this dissertation also makes a contribution to the education and developmental science research literature on curriculum-based interventions. The findings of the PCER study were largely null, although several analytic issues, such as low statistical power, have been cited to explain the lack of significant effects (PCER, 2008). My re-analysis of the data show that there were heterogenous treatment effects previously hidden. By conducting the baseline subgroup analyses as well as the quantile treatment effects estimation, I am able to compare the findings obtained by the two methods. I find that while the baseline subgroup analyses often provide an accurate summary of differences across the distribution, they might also miss important variation in differences. Though differences between subgroups are

important and tell us about how these groups differ on average, differences at the extremes—the top and bottom of the distribution—often vary substantially. In cases such as *skill-specific* curricula effects on differently *skilled* children, where the extremes of the distribution are of particular interest, quantile treatment effects estimation is a useful method to employ for examining heterogeneity of treatment impacts. This method also suggests that children at different parts of the ex-post distribution of literacy and language achievement are not subject to identical processes during the intervention. Given the growing interest in understanding treatment effect heterogeneity and the extent to which interventions work for some children and not others (Duncan & Vandell, 2011), researchers are encouraged to utilize various novel methods to understand variation in treatment impacts.

Future Research

Strong empirical research on early childhood curricula is a top priority and there is still much left to be learned about various aspects of curricula. A more in-depth understanding of the content of the whole-child curricula under investigation is needed in order to explain how they are distinct from one another and whether they produce any positive impacts on the classroom environment or children's outcomes. A systematic content analysis can help identify key components of curricula that may offer a new lens through which classroom practices can be characterized and potentially linked to children's school readiness. By characterizing curricula in distinguishable ways, future research can move beyond the often-examined granular teacher instructional domains of standard observational tools. This content analysis is also a fruitful area of research with direct policy implications as it can provide specific detail on content and modes of instruction, which is intended to inform decisions on which curricular package to purchase that best suits the unique needs of public preschool programs.

Looking beyond individual curricula, one promising approach to promoting children's school readiness is through a hybrid curriculum model that blends together a rich academic curricular approach with playful learning and then focus on ensuring that it is implemented in classrooms as faithfully as possible (Fuller, Bein, Bridges, Kim, & Rabe-Hesketh, 2017; Jenkins & Duncan, 2017). The Boston Pre-K program is one example of an integrated playful learning approach that incorporates the whole-child philosophy while also focusing on specific learning outcomes (Weiland & Yoshikawa, 2013). These integrated curriculum approaches appear to be quite promising, as it allows preschool teachers to focus on improving children's readiness for school across multiple domains—such as language, literacy, mathematics, and social-emotional skills, and not just language skills. Such models need to be evaluated in future research using a strong design that can track impacts on child outcomes during and beyond elementary school. Most importantly, it needs to be shown to be replicable at scale in other school systems serving predominantly low-income children.

This dissertation aimed to answer an important overarching question that is my larger research agenda: What makes preschool effective for children? Curricula set the goals for the knowledge and skills that children should be able to acquire in early learning setting (Jenkins & Duncan, 2017). A strong, well-defined curriculum is arguably one of many "active" ingredients that makes preschool effective for children. Weiland and colleagues (Weiland, McCormick, Mattera, Maier, & Morris, 2018; Yoshikawa et al., 2013; Yoshikawa, Weiland, & Brooks-Gunn, 2016) point to the "strongest hope" model as a promising tool for studying these mechanisms of preschool program effectiveness, including intentionally-focused content-specific curricula, teacher training, and coaching. This dissertation was motivated by this first element of their model and Chapter 3 alluded to the last two, but it will be important for future research to

pinpoint how all three interact together to bring about educationally and statistically significant improvements across multiple domains of children's school readiness.

Conclusion

Social, economic, and political conditions place mounting pressures on early childhood programs to produce increases in child outcomes and to justify public and private investments in the maintenance and expansion of early education and care. This dissertation and my meta-analytic work both show that curricula are deeply important for shaping classroom experiences, that selection and implementation of curriculum can have an impact for better or worse on children's learning and development, and that there is a critical need for evidence-based early childhood curricula. Most importantly, this work points to ways that curricula can best be leveraged to improve outcomes for the children that preschool programs serve.

References

Clements, D. H. (2007). Curriculum Research: Toward a Framework for "Research-Based Curricula." *Journal for Research in Mathematics Education*, 35-70.

Clifford, R. M., Barbarin, O., Chang, F., Early, D. M., Bryant, D., Howes, C., et al. (2005). What is Pre-Kindergarten? Characteristics of Public Pre-Kindergarten Programs. *Applied Developmental Science, 9*(3), 126-143.

Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation*, *11*(2), 193-221.

Durlak, J. A. (2010). The importance of doing well in whatever you do: A commentary on the special section, "Implementation research in early childhood education." *Early Childhood Research Quarterly, 25*, 348–357.

Durlak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.

Dusenbury, L., Brannigan, R., Hansen, W. B., Walsh, J., & Falco, M. (2005). Quality of implementation: Developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research*, *20*(3), 308-313.

Frede, E., & Ackerman, D. J. (2007). *Curriculum decision-making: Dimensions to consider.* New Brunswick, NJ: National Institute for Early Education Research.

Fuller, B., Bein, E., Bridges, M., Kim, Y., & Rabe-Hesketh, S. (2017). Do academic preschools yield stronger benefits? Cognitive emphasis, dosage, and early learning. *Journal of Applied Developmental Psychology*, *52*, 1-11.

Gilliam, W. S., Ripple, C. H., Zigler, E. F., & Leiter, V. (2000). Evaluating child and family demonstration initiatives: Lessons from the Comprehensive Child Development Program. *Early Childhood Research Quarterly*, *15*(1), 41-59.

Jenkins, J. M., & Duncan, G. J. (2017). Do pre-kindergarten curricula matter? In D. Phillips, K. A. Dodge, & Pre-Kindergarten Task Force (Eds.), The Current State of Scientific Knowledge on Pre- Kindergarten Effects (pp. 37-44). Washington, DC: Brookings Institution and Duke University.

Jenkins, J.M., Duncan, G.J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. (2018). Boosting school readiness: Should preschools target skills or the whole child? *Economics of Education Review*. doi: 10.1016/j.econedurev.2018.05.001

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, *23*(1), 51-68.

Lipsky, M. (1980). *Street Level Bureaucrats.* New York, NY: Russell Sage Foundation.

McCall, R. B. (2009). Evidence-based programming in the context of practice and policy. *Social Policy Report*, *23*(3), 3-18.

Nguyen, T. (2017). Impacts of Targeted and Global Preschool Curricula on Children's School Readiness: A Meta-Analytic Review. Working Paper, Irvine Network on Interventions in Development.

Preschool Curriculum Evaluation Research Consortium (PCER). (2008). *Effects of preschool curriculum programs on school readiness*. (NCER 2008-2009). National Center for Education Research, Institute for Education Sciences, U.S. Department of Education, Washington, DC.

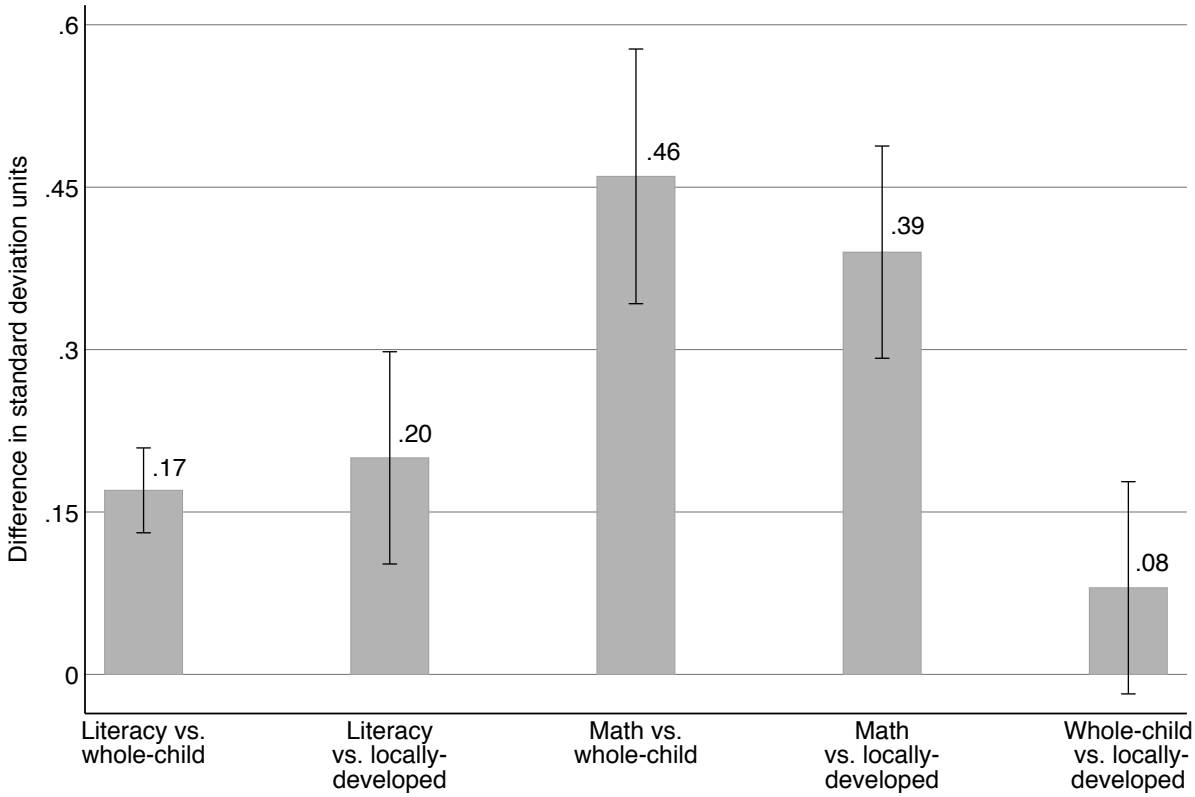
Weiland, C., McCormick, M., Mattera, S., Maier, M., & Morris, P. (2018). Preschool Curricula and Professional Development Features for Getting to High-Quality Implementation at Scale: A Comparative Review Across Five Trials. *AERA Open*, *4*(1), 1-16. doi: 10.1177/2332858418757735

Weiland, C., & Yoshikawa, H. (2013). Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills. *Child Development, 84*(6), 2112-213.

Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *The Future of Children, 26*(2), 21-35.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W., et al. (2013). Investing in our future: The evidence base on preschool education. New York, NY: Foundation for Child Development, Society for Research in Child Development.

Figure 5.1. Impacts of various curricula on academic outcomes from Nguyen (2017) meta-analysis.



Note. The height of each column represents the covariate adjusted mean of effect size estimates within each group, controlling for study characteristics and weighted by the inverse of the standard error squared of the individual effect size estimate. Bars show 95% confidence interval. The first and second bars are effect size estimates for literacy and language outcomes. The third and fourth bars are effect size estimates for math outcomes. The last bar represents effect size for composite scores of literacy and language outcomes and math outcomes. Chapters 2 and 3 of this dissertation focuses on the last bar, which includes a subset of whole-child curricula that were included in the calculation of that effect size. Chapter 4 of this dissertation focuses on the first and second bars, drawing upon data from a subset of curriculum interventions that were included in the calculation of those effect sizes.