ACCESS, QUALITY, AND STUDENT OUTCOMES IN A CHANGING EARLY CHILDHOOD LANDSCAPE

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Scott Latham, B.A, M.P.P.

August, 2016

Educational Policy Studies
Curry School of Education
University of Virginia
Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, *Access, quality, and student outcomes in a changing early childhood landscape*, has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy

_____
Daphna Bassok, Chair

_____
Jim Wyckoff

_____
Vivian Wong

_____
Robert Pianta

_____
Thomas Dee (Stanford)

June 21, 2016

TABLE OF CONTENTS

Dedication…….....iv
Acknowledgements……....v
List of tables……..vi
List of figures……..viii

Linking document……...1

Chapter 1 – Kids Today: Changes in school readiness in an early childhood era…….6

Chapter 2 – The effects of accountability incentives in early childhood education…….46

Chapter 3 – Examining the gap in preschool enrollment between Hispanic and non-Hispanic children……..105

This dissertation is dedicated to Jay and Terri Latham. There are many great parents in the world, but I wouldn't trade mine for any of them.

ACKNOWLEDGEMENTS

Daphna Bassok, my adviser, mentor, and advocate. I am eternally grateful for your brilliance and wisdom, and this dissertation would not be half of what it is without your help.

Jim Wyckoff, thank you for taking a chance on me five years ago, it's hard to imagine now that I could have ended up anywhere other than UVA. Thanks for establishing and constantly cultivating a culture of collaboration, support, and intellectual rigor.

My soon-to-be wife Emily Handy, you were an endless source of support and you somehow kept my life on track even when I was helpless to do anything but dissertate.

Mindy Adnot and Veronica Katz, I'm very lucky to have belonged to a cohort like ours. You were some of my best friends, allies, and sounding boards.

Marc Latham and Jenna Diorio, my brother and sister, you both deserve a shout out because I love you each to death, but you didn't help much with the dissertation :).

And to my parents again, you were always loving, caring, and supportive, and you put up with my incessant questions. And now I'm a professional question asker.

LIST OF TABLES

Table		Page
1.1	Differences in teacher-rated academic and behavioral outcomes, across cohorts (OLS estimates)	36
1.2	Differences in teacher-rated math and literacy proficiency across cohorts, by race and SES (OLS estimates)	37
A1A	Correlations between teacher-reported measures and direct student assessments	40
A1B	Covariate descriptive statistics	41
A1E	Differences in teacher-rated student proficiency, across cohorts (bounded OLS estimates)	45
2.1	Descriptive statistics for the analytic sample at baseline (T) through T+5	95
2.2	First stage estimates across specifications and bandwidth restrictions	96
2.3	Auxiliary RD regressions of baseline covariate balance in terms of auspice	97
2.4	Reduced-form RD estimates for outcomes at baseline (T) through T+5	98
2.5	Heterogeneity of results by enrollment and auspice	99
A2C	Comparison of sample and nonsample providers 2007-2009	103
A2D	Sensitivity of results in year T+5 to alternate specifications and bandwidths	104
3.1	Child care use among children 0-5, by race/ethnicity	149
3.2	Differences in family predictors of early child care use, by race/ethnicity	150
3.3	Odds of Hispanic and black children 0-5 attending primarily center-based care, relative to white children	152

3.4 Child care use among Hispanic children 0-2, by home language and immigrant status……153

3.5 Child care use among Hispanic children 3-5, by home language and immigrant status……154

3.6 Odds of Hispanic children primarily attending center based care relative to white children, among children 0-2……155

3.7 Odds of Hispanic children primarily attending center-based care relative to white children, among children 3-5, not yet in K……156

A3A Characteristics of Hispanic households by home language and immigrant status……157

LIST OF FIGURES

Figure		Page
1.1	Distribution of selected teacher-reported literacy and math skills	38
1.2	Percentage of students who were low and high proficiency in math across cohorts	39
A1C	Distribution of additional teacher-reported math and literacy skills	43
A1D	Percentage of students who were low and high proficiency in literacy across cohorts	44
2.1	First stage relationships between average ERS ratings and star ratings at baseline	89
2.2	Density of forcing variable around RD threshold	90
2.3	Reduced form relationships between average ERS ratings at baseline (T) and star ratings in subsequent years	91
2.4	Reduced form relationships between average ERS ratings at baseline and in T+5	92
2.5	Reduced form relationship between average ERS ratings at baseline and enrollment/closures	93
A2A	Calculation of program standards scores in North Carolina	100
A2B	Sample five star rated license	102

LINKING DOCUMENT

A large and consistent body of research shows that early childhood is a critical time for student learning. In particular, research from economics, sociology, and psychology has shown that high quality early child care settings, such as preschool and other center-based settings, have the potential to improve short- and long-term outcomes and change academic trajectories (Camilli, Vargas, Ryan, & Barnett, 2010; Campbell et al., 2012; Deming, 2009; Schweinhart et al., 2005; Weiland & Yoshikawa, 2013). Based on the promise of this research, policymakers have become increasingly focused on investing in programs that serve children before they begin kindergarten. For instance, state spending on preschool more than doubled from 2001-2015, with over $6.2 billion spent on state-funded preschools in the most recent year alone.

As research continues to show that *high quality* care is necessary to produce the strong benefits of preschool programs, states have increasingly adopted accountability systems in early child care markets to improve the quality of providers. These investments are coming fast but often without empirical support. At the same time, instruction in early childhood has become increasingly focused on academic skills such as literacy and math (Bassok, Latham, & Rorem, 2016). Amidst all these changes to the early childhood landscape, there are many unanswered questions that are relevant for policymakers. In particular, have all of these changes led to differences in student learning? Have they led to more equal access to preschool in general, or to high quality preschool? Have accountability systems increased the quality of child care providers? This dissertation aims to provide new insights toward these questions, leveraging four

separate datasets to answer both descriptive and causal questions. I tackle a diverse set of topics, using both quasi-experimental and descriptive analyses, and leverage four separate datasets

In chapter 1 (coauthored with Daphna Bassok), we explore how young children's teacher-reported math, literacy, and behavioral skills at kindergarten entry have changed across the years 1998 to 2010. Despite an influx of funding into early childhood programs (Barnett et al., 2016), and an increased focus on academics in early childhood (Bassok, Latham. & Rorem, 2016) we have little empirical evidence on whether children in recent cohorts are arriving in kindergarten with a different set of skills than in previous years. In particular, we might expect that entering kindergarteners in recent years will have had more exposure to academic content such as literacy and math, and will thus be more proficient. This paper addresses this question directly, using two large, nationally representative datasets of children entering kindergarten in 1998 and 2010. We find that students in the more recent cohort entered kindergarten substantially more proficient across teacher-reported literacy and math skills. We also find that improvements over this period were disproportionately large among black children, with marginal evidence of disproportionate effects among Hispanic children. These results suggest that early achievement gaps between white and black/Hispanic children may be narrowing, and are consistent with other work that documented changes in achievement gaps over the same period (Reardon & Portilla, 2016). Finally, we find that students in the more recent cohort had *worse* behavioral outcomes in terms of self control and approaches to learning (e.g. eagerness to learn; ability to pay attention), a potentially troubling finding given

evidence that these attention measures are associated with later school achievement (Duncan, 2007).

This paper provides a nationally representative exploration of whether levels of school readiness have changed over a period of heightened investment in early childhood education and provides compelling evidence that they have. The results are encouraging, and consistent with other recent work that finds narrowing of early achievement gaps in recent years (Reardon & Portilla, 2011). More work is needed to explore the causes of these changes in student skills over time. Future work should also explore whether gains in early math and reading skills are maintained as children age.

Chapter 2 (coauthored with Daphna Bassok and Thomas Dee) explores the effects of an early childhood accountability system in North Carolina. In particular, accountability systems (called Quality Rating and Improvement Systems, or QRIS) have been implemented throughout the nation, with 39 states adopting a QRIS as of early 2015, and nearly all others in the planning or piloting phase. Despite substantial investment in these early childhood accountability efforts, there is virtually no evidence on the extent to which such initiatives can lead either to improvements in program quality or to changes in parents' decisions about their children's early learning experiences.

We provide evidence on this high-profile policy initiative by examining North Carolina's Star Rated License program, one of the oldest and most well established QRIS in the country. We implement a regression discontinuity design, leveraging the fact that small differences in an observed measure of quality led to large "jumps" in the probability of receiving a lower versus higher quality rating. Specifically, we examine

whether and how providers responded to being quasi-randomly assigned to a lower quality rating by examining future indicators of provider quality. We also examine whether parents responded to information about quality by examining changes in enrollment.

We find that providers responded in multiple ways. In particular, some responded by increasing their score on an observational measure of quality. By contrast, others responded by opting not to be re-rated on that same observational measure. We also find that centers who received lower ratings had significantly lower enrollment five years after the initial ratings occurred. This finding provides suggestive evidence that parents may have responded to information about quality by changing their enrollment decisions. This work provides the first credibly causal evidence on a critical component of the theory of change that underlies QRIS programs, and finds evidence that fully-implemented, scaled-up accountability systems can work through two of its primary theorized mechanisms.

In my final chapter I explore differences in preschool enrollment between Hispanic and non-Hispanic children. According to recent estimates, 37% of 3 and 4 year old Hispanic children attended some kind of preschool, compared with 51% of white children of the same age (Coley et al., 2014). This "gap" in preschool enrollment has been documented for decades, and likely contributes to the early achievement gap between Hispanic and non-Hispanic children. Although research has attempted to explain the reasons for the difference in enrollment, we still have an under-developed understanding of the phenomenon. In particular, research that has explored these

questions uses decades-old data, and generally documents patterns among preschool age children (3-5), but not younger children. Further, research has not adequately described how patterns differ *within* Hispanic children (e.g. by home language or by immigrant status). This paper fills this gap, leveraging unique, newly-available national data, to address these questions, and exploring the extent to which differences in enrollment are associated with 1) income/SES, 2) household composition, and 3) parent work schedules.

In keeping with previous work, I find that Hispanic children are less likely to enroll in preschool than non-Hispanic white or black children. However, this differs across groups of Hispanic children. Those from households that speak only English were similar to white children in their likelihood of attending preschool. By contrast, children from Spanish-speaking and immigrant households were considerably less likely to attend, though it is important to stress that these families are also of much lower income on average. Indeed, differences in income and socioeconomic status accounted for most of the difference in preschool enrollment. These descriptive findings highlight the heterogeneity in preschool utilization patterns of Hispanic families. Future work should attempt to identify causal impact of policies aimed at changing preschool enrollment. Future work should also examine additional factors that may affect parent child care decisions such as parental preferences and availability of child care.

Taken together, these papers provide new policy insights in the field of early childhood education. Each explores timely and unanswered policy questions in a period of heightened attention and investment paid to early childhood, and each provides a unique contribution to the existing literature.

Chapter 1 - Kids today: Changes in school readiness in an early childhood era

Daphna Bassok & Scott Latham

Abstract

Public investment in early childhood education has expanded dramatically in recent years. Despite this investment, we have little empirical evidence on whether children today enter school with different skills than they did in the late nineties. Using two large, nationally representative datasets, this paper documents how students entering kindergarten in 2010 compare to those who entered in 1998 in terms of their teacher-reported math, literacy and behavioral skills. Our results indicate that students in the more recent cohort entered kindergarten substantially more proficient at both math and literacy skills. Increases in academic skills over this period were particularly pronounced among black children. Implications for policy are discussed.

Over the past two decades, public investment in early childhood education has grown rapidly. Between 2001 and 2014 state spending on preschool initiatives more than doubled from $2.4 to $5.6 billion and since the early nineties the number of children in public preschool has also nearly doubled (Barnett, Carolan, Squires, Clarke Brown, & Horowitz, 2015; U.S. Census Bureau, 2015). Many states have recently introduced early learning standards, more restrictive quality regulations for early childhood education providers, and Quality Rating and Improvement Systems (QRIS), accountability systems that incentivize quality improvements in early childhood settings (The Build Initiative & Child Trends, 2015).

Despite the unprecedented interest and investment in early education, we have little empirical evidence on whether children entering kindergarten in recent years have stronger math and literacy skills at school entry than they did in the past, and relatedly whether their behavioral skills have changed. Although datasets like the National Assessment of Educational Progress (NAEP) have long allowed for comparisons over time in the academic achievement of elementary, middle and high school aged students, until now we have had no national data that would allow for comparisons across cohorts of children at kindergarten entry. The current paper aims to fill this gap.

Using two nationally representative samples of children entering kindergarten in 1998 and 2010 this paper asks three related research questions:

1. To what extent do children who entered kindergarten in 2010 differ from those who entered school in 1998 with respect to teacher-reported measures of math, literacy & behavior?

2. Do changes in demographic characteristics, preschool participation and other observable factors over this period explain observed changes in school readiness?

3. Do changes in school readiness over time differ across racial and socio-economic subgroups leading to changes in school-entry achievement gaps?

Background

Some research indicates that children's early academic skills are strongly predictive of outcomes well into the future, including college attendance, home ownership, earnings, and retirement savings (Chetty et al., 2011). It is also well documented that by kindergarten entry there are large achievement gaps based on race and income, and that these gaps persist as children proceed through school (Fryer & Levitt, 2004; Reardon, 2011).

Over the past three decades a large body of evidence from education, developmental psychology, neuroscience, and economics has demonstrated that early childhood is a particularly malleable time in the life course and that interventions targeted towards this period can have both long-lasting and cost-effective impacts (Bassok & Loeb, 2015; Heckman, 2006; Shonkoff & Phillips, 2000). Bioecological theory suggests that optimal development occurs when children experience consistent and supportive interactions with the people and objects in their immediate environment (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 2006). By providing a stimulating and enriching environment for children in the years before kindergarten, early interventions can potentially prevent gaps from developing, or mitigate their severity. Indeed, a large body of evidence suggests early childhood programs can have important short- and long-term benefits

(Camilli, Vargas, Ryan, & Barnett, 2010; Campbell et al., 2012; Deming, 2009; Schweinhart et al., 2005; Weiland & Yoshikawa, 2013).

Growing understanding of both the importance of early childhood in the life course *and* the documented benefits of high-quality preschool interventions has led to a sharp increase in public support for early childhood opportunities. One plausible, though untested, hypothesis is that the rapid expansion of public preschool since the late nineties has led to an improvement in children's "school readiness" broadly defined.

The limited existing evidence does support the notion that children are entering schools more "ready" than they once were. Using repeated, nationally representative surveys of parents with children ages 3-6, a recent report showed that between 1993 and 2012, the percentage of parents who said their child could recognize all the letters in the alphabet nearly doubled from 21 to 38 percent (Child Trends, 2015). Similarly, the percentage of parents that indicated their children could count to 20 increased by 16 percentage points. Although that study relied solely on parent reports of four relatively crude measures of children's school readiness skills, the findings suggest that young children acquire basic academic skills at an earlier age than they once did.

Similarly, three states that have been conducting school readiness assessments for over a decade recently released reports that indicate children are increasingly arriving at kindergarten "ready to learn" based on assessments at school entry (Maryland State Department of Education, 2014; Minnesota Department of Education, 2013; Virginia Performs, 2015).

While these trends are aligned with our hypothesis, to date there has not been a systematic empirical study documenting national changes in school readiness over time. This is the primary goal of the current study. In addition, we examine whether changes in school readiness over time have differed across groups of children. Because the majority of public early childhood programs are targeted towards low-income students (Barnett et al 2015) we might expect that improvements in children's academic skills at kindergarten entry would be particularly pronounced among low-income and minority children. Consistent with this hypothesis, recent work by Reardon & Portilla (2015) demonstrates that between 1998 and 2010 racial and income-based gaps in reading and math scores at school entry have narrowed. Note, however, that the narrowing achievement *gaps* documented tell us nothing about changes in the *levels* of academic skills at school entry (i.e. gaps could narrow even if average school readiness skills have dropped or stayed the same). In the current study we explore whether non-white and poor children experienced greater changes in academic outcomes than did their peers.

Finally, the current study adds to the existing literature by documenting national trends in children's behavioral skills at school entry. It is difficult to provide directional hypotheses about changes over time in children's behavior. On one hand, if more children are experiencing preschool in the year prior to kindergarten, perhaps they enter kindergarten more prepared to meet classroom norms and behavioral expectations. On the other hand, research demonstrates that preschool participation is negatively associated behavioral outcomes in subsequent years (Belsky et al., 2007; Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007). Researchers

have also hypothesized that exposure to structured literacy or math curriculum in early childhood may be associated with stress, problem behaviors or other unintended behavioral consequences (Christakis, 2016; Stipek et al., 1998; Stipek, Feiler, Daniels, & Milburn, 1995). We provide descriptive evidence to address these hypotheses.

Method

Data

The National Center for Education Statistics (NCES) has tracked two large, nationally representative cohorts of children longitudinally starting in kindergarten through its Early Childhood Longitudinal Study (ECLS) program. The first cohort began kindergarten in 1998 and the second started in 2010.[1] Because the studies provide largely overlapping and comparable measures (West, Denton, & Germino-Hausken, 2000; West, Denton, & Reaney, 2001), the combined datasets provide a unique opportunity to assess whether there have been nationwide changes in school readiness, both for the overall population and for specific subgroups, over a period characterized by heightened investment in early childhood.

Each dataset was collected using a multi-stage probability sample design; children were selected from schools which were first selected from "primary sampling units" (counties or groups of counties). Schools and primary sampling units were both chosen with probability proportional to size. Both datasets include

[1] The more recent ECLS-K cohort began collecting data from students in the fall of 2010 but is officially referred to as the ECLS-K 2011. This paper, however, uses data collected in the fall, so the years of the study are more accurately reported as 1998 and 2010

direct student assessments of children as well as parent and teacher surveys in the fall and spring of the kindergarten year, and a school administrator survey in the spring (Tourangeau, Nord, Lê, Sorongon, & Najarian, 2009). The original ECLS-K followed students through 8th grade, and the 2010 cohort will be followed through fifth grade. We limit our analysis to first-time kindergarteners in each cohort. We limit our sample to observations with non-missing data for all outcomes considered, and construct two separate samples when considering academic and behavioral outcomes. The academic sample includes approximately 16,050 students in 1998 and 13,500 in 2010, and the behavioral sample includes approximately 14,750 and 11,900 students in 1998 and 2010 respectively (all sample sizes rounded to the nearest 50 in accordance with NCES guidelines).

We conduct multiple imputation using chained equations to avoid the bias that may arise when analyzing complete-case data. Our imputation model accounts for all the covariates that we later include in our analysis (i.e. demographics, home environment, and kindergarten teacher characteristics), and we impute independent but not dependent variables. Multiple imputation was conducted using the MI command in Stata, and following Allison (2012) 20 imputed datasets were generated.

Measures of school readiness. The ECLS-K datasets contain multiple potential measures of children's school readiness including direct assessments and teacher-reported measures of early literacy and mathematics knowledge, as well as teacher-reported measures of children's behavior.

Teacher-reported measures. Ideally, to analyze changes in students' math and literacy ability over time we would compare direct assessments across the two cohorts. Although both ECLS-K waves directly assessed students' math and literacy ability, these assessments are not directly comparable across the two cohorts (Tourangeau et al., 2013). As a result, this analysis relies on teacher-reported measures of student skills at kindergarten entry.

Existing research demonstrates that teachers' assessments of students' cognitive skills are strongly correlated with both current and future direct assessments (Hecht & Greenfield, 2001; Hoge & Coladarci, 1989; Teisl, Mazzocco, & Myers, 2001). This pattern certainly holds in both waves of the ECLS-K as shown in Appendix 1A. The table also highlights that for the earlier cohort, teacher-reported measures of children's early readiness skills are predictive of direct assessments not only in kindergarten but all the way through eighth grade.

Despite the strong correlation between teacher reports and direct assessments, caution is warranted when using teacher assessments to measure children's "true" ability. Earlier studies have shown that a portion of the variation in teacher assessments of young children is explained by both teacher characteristics (e.g. education levels, experience) and child characteristics (e.g. race, socio-economic status (SES)) (Kilday, Kinzie, Mashburn, & Whittaker, 2012; Mashburn, Hamre, Downer, & Pianta, 2006). Despite this, teacher assessments are the most widely-used, cost-effective and efficient method for assessing young children. In fact, teachers' extensive interactions with children may allow them unique insights

and knowledge not captured by direct assessments (Epstein, Schweinhart, DeBruin-Parecki, & Robin, 2004).

A further concern when using teacher assessments to track *changes* in children's skills over time is that observed changes may be driven in part by changes over time in how teachers perceive children or in teacher's approaches to assessing children. We have no way to definitively assess whether teachers have changed the way they assign ratings over time, an important limitation we return to at the end of the paper.

***Academic outcomes*.** Both ECLS-K datasets include teacher-reported measures of student proficiency across a broad range of math and literacy skills. In the first months of kindergarten (September-December), teachers were asked to rate each child's proficiency in the following 14 domains on a scale from 1 ("Child has not yet demonstrated skill") to 5 ("Child demonstrates skill competently and consistently"):[2]

[2] We omit students from our analysis if their teachers reported that topics had not yet been introduced in their classrooms because teachers in these cases have not had a chance to assess student proficiency. However, this exclusion could bias our results if these types of students generally were less proficient than students for whom teachers provided ratings. We performed a bounding exercise to explore the extent to which this could be driving our results. For the purposes of this exercise, we assumed that all students who were missing data for a given skill were not proficient at that skill (i.e. that they were the lowest level of proficiency). We then re-ran our analysis and present results in Appendix 1E. The results for literacy are quite similar to our main results, suggesting that these results are unlikely to suffer from this type of bias. Estimates for differences in math across time are about 60% as large as our main results. Notably, even using this extremely conservative assumption, we find meaningful differences over time in student reading and math abilities.

Math skills

- Sorts math materials by various rules and attributes
- Orders groups of objects (by height, color, etc.)
- Understands relative quantities
- Solves problems using numbers
- Understands graphing activities
- Uses instruments accurately for measuring
- Uses a variety of strategies to solve math problems

Language and literacy skills ("Literacy")

- Uses complex sentence structures
- Understands and interprets stories read to him/her
- Easily names all upper and lower case letters
- Predicts what will happen next in stories
- Reads simple books independently
- Demonstrates early writing behaviors
- Understands conventions of print

We analyze changes in proficiency over time for each individual skill, and also construct three summary measures for each subject. First, to construct measures of students' overall proficiency in math and literacy, we average across the items within each subject (e.g. the average of 7 math skills is the "overall math" proficiency). Next, for each subject we construct indicator variables labeled "low proficiency" and "high proficiency" to indicate whether students were in the tails of the distribution. Specifically, we define "low proficiency" to mean that a student was rated either a 1 or 2 on at least half of the skills considered, and "high proficiency" to mean that a student was rated either a 4 or 5 on at least half of the skills. For example a student who was rated a 1 or 2 on at least 4 of 7 math skills would be classified as having low math proficiency.

Because teacher assessments were collected during the first few months of kindergarten, they are not "pure" measures of student knowledge at school entry

and may be capturing, in part, skills gained during the beginning of the school year. This concern is lessened because the data collection period is extremely similar across the two studies. All of our estimates also control for the amount of time that children spent in kindergarten before assessment.

***Behavioral outcomes*.** In the fall of the kindergarten year, teachers completed a version of the Social Skills Rating System (Gresham & Elliott, 1990), a widely used assessment of social and emotional development, and a scale that measures student approaches to learning. Respondents were asked to rate the frequency of different types of student behavior on a scale from 1 ("Never exhibits behavior") to 4 ("Exhibits behavior frequently"). These items were then combined into "subscales" which capture different dimensions of student behavior. The ECLS-K contains information about five subscales, which we report along with both the number of items and the subscale reliability coefficients. Reliability coefficients were quite similar across cohorts, so we report pooled coefficients here for all subscales except one (approaches to learning), which differed slightly in the number of items across cohorts. The five subscales were self-control (4 items, reliably coefficient (RC) = .85), interpersonal skills (5 items, RC = .88), externalizing problem behavior (5 items, RC = .89), internalizing problem behavior (4 items, RC = .80) and approaches to learning (6 items in 1998, RC = .89; 7 items in 2010, RC = .91).

These behavioral measures have skewed distributions, in that teachers report kindergarteners are generally well-behaved. Following Grimm et al. (2010), we address this issue by dichotomizing behavioral measures to construct indicators of problem behavior. Three of the measures considered indicate positive behavior

(i.e. self-control, interpersonal behavior, approaches to learning) and so for these we construct indicators for whether a student was at least 1 SD *below* the 1998 mean. By contrast, for the two outcomes indicating negative behavior (i.e. internalizing and externalizing) we construct indicators for whether a student was at least 1 SD *above* the 1998 mean.

Early childhood experiences. To measure access and exposure to early childhood education we use parent reports of the type of care their children received in the year prior to kindergarten. Specifically, we account for whether a child attended "formal" care (defined as either Head Start, pre-kindergarten or center-based care), whether a child attended a *public* formal care program, the number of hours a child spent in formal care each week, and whether a child attended kindergarten and prekindergarten in the same building, a proxy for school-based prekindergarten, which is associated with greater academic benefits for children (Magnuson, Ruhm, & Waldfogel, 2007). We also consider principal reports of whether the child's kindergarten school also offers prekindergarten. Appendix 1B provides descriptive statistics for these and all other variables included in our analysis, separately by cohort. We do not observe meaningful changes in formal care utilization; about two thirds of children attended formal care in both waves. However, consistent with heightened investments in public preschool, we note substantial increases in publicly-funded preschool and in elementary schools offering kindergarten programs.

Additional covariates. We disaggregate our analysis by race and a measure of SES constructed from parental income, education, and occupational prestige, which we divide into quintiles.

Although this manuscript is motivated by the rapid expansion of public early childhood education opportunities, many other factors changed over the study period, and in a final set of analyses we assess whether changes in three other sets of covariates explain changes in school readiness: demographic characteristics of children and families; children's home learning environment and parental interactions; and teacher characteristics. A key strength of the ECLS-K datasets is that they provide unusually rich data about these measures.

Because demographic changes may be associated with changes over time in children's school readiness, we account for age (both at kindergarten entry and at assessment), gender, whether children were born in the U.S., whether they are U.S. citizens, and whether English is the primary language (or spoken at all) in children's homes. We also account for the region of the country in which children reside. Appendix 1B highlights changes in these variables over time; most notable is the increase in the percentage of Hispanic kindergarteners. In a recent study, Bassok et al. (2016) use the same ECLS datasets leveraged in the current study to examine changes over time in early childhood experiences . That analysis showed substantial increases in the time parents spent engaging with their children, the amount of learning materials children had in the home (e.g. books, educational computer games), and the expectations parents held about school readiness. Further, many of these measures increased disproportionately among low-income children.

These changes may have had meaningful implications for children's school readiness and to address this we include a host of measures of children's home environments. Parents were asked about the frequency with which they do activities with their children, such as reading books, playing games, and doing chores. Parents were also asked to rate the extent to which they think a number of skills are important for school readiness. These skills include counting to 20, knowing the letters of the alphabet, sharing with others, using a pencil, paying attention/sitting still, and communicating needs/wants verbally. Parents were also asked a series of questions regarding their child's use of computers in the home. They were asked whether their child uses a computer, how frequently, and whether the child uses the computer to access the internet and for educational purposes. Appendix 1B highlights systematic increases across many of these measures.

Finally, kindergarten teachers reported detailed information about their demographics, teaching experience, and education. We include these measures because earlier studies have highlighted the association between teacher characteristics and their assessments of child outcomes in early childhood settings (Mashburn et al., 2006).

Analysis

We compare measures of school readiness over time using Ordinary Least Squares (OLS) (and Linear Probability Models for the dichotomous outcome variables).[3] To describe changes over time in these measures, we specify the following model:

[3] Some of our outcomes are dichotomous and are thus most appropriately modeled using a limited dependent variable specification. We have estimated logit models for all dichotomous outcomes considered in this paper and the results are strikingly

$$y_i = \beta_0 + \beta_1 ECLS2010_i + \beta_2 Age_i + \varepsilon_i \quad (0)$$

Here, y_i refers to either an academic or behavioral outcome for student *i*, and $ECLS2010_i$ is an indicator set to 1 if student *i* was part of the 2010 ECLS-K cohort, and set to 0 if the student was part of the 1998 cohort. Age_i represents student *i*'s age in months both at kindergarten entry and at the time of assessment. This controls for potential differences in age across cohorts that may be confounded with differences in cognitive and social abilities, as well as differences in timing of assessments across cohorts. ε_i represents an error term with mean 0. β_1 is the coefficient of interest, and it provides an estimate of the (age-adjusted) difference across cohorts for each outcome variable. We also employ probability weights that adjust for non-response, making the results nationally representative. Standard errors are clustered at the teacher level.

After describing the raw magnitude of the changes over time, we explore how the results change when accounting for changes in demographics across this time. Here, we estimate the following model:

$$y_i = \beta_0 + \beta_1 ECLS2010_i + \beta_2 Age_i + \beta_3 Demographics_i' + \varepsilon_i \quad (1)$$

where $Demographics_i'$ is a vector that includes race, SES, ELL status, and whether the child is U.S. born or a U.S. citizen. We use β_1 from model (1) as the starting point from which we will try to account for differences across cohorts using observable measures of preschool attendance, the home environment, and kindergarten teacher characteristics. Here, we estimate the following series of models:

---

similar to those from the OLS models. As a result, we present OLS estimates to facilitate easier interpretation.

$$y_i = \beta_0 + \beta_1 ECLS2010_i + \beta_2 Age_i + \beta_3 Demographics_i' + \beta_4 Preschool_i' + \varepsilon_i \quad (2)$$

$$y_i = \beta_0 + \beta_1 ECLS2010_i + \beta_2 Age_i + \beta_3 Demographics_i' + \beta_4 Preschool_i' + \beta_5 Home\_env_i + \beta_6 K\_tch\_chars_i' + \varepsilon_i \quad (3)$$

Here, *Preschool$_i$'* is a vector that includes variables relating to preschool care in the year before kindergarten. *Home_env$_i$'* is a vector that includes parent beliefs about the importance of different skills for kindergarten readiness, activities in which the child participates, and measures of computer availability and use. *K_tch_chars$_i$'* includes information about kindergarten teachers' background and qualifications. To the extent that we find differences in β_1 between model (1) and models (2) or (3), this would suggest that changes in student outcomes were explained, at least partly, by the included covariates. Finally, we examine whether differences in student outcomes across cohorts vary by race and socioeconomic status. Here, we estimate models of the form:

$$y_i = \beta_0 + \beta_1 ECLS2010_i + \beta_2 Age_i + \beta_3 Race_i' + \beta_4 ECLS2010_i * Race_i' + \varepsilon_i \quad (4)$$

$$y_i = \beta_0 + \beta_1 ECLS2010_i + \beta_2 Age_i + \beta_3 SESQ1_i + \beta_4 ECLS2010_i * SESQ1_i + \varepsilon_i \quad (5)$$

In model (4), *Race$_i$'* includes indicators for whether a student is black, Hispanic, Asian, or other nonwhite (omitting white). These indicators are then each interacted with the 2010 cohort to explore whether there were disproportionate changes over time for students of differing race/ethnicity. In model (5), we include an indicator for whether a student was in the lowest SES quintile (i.e. the poorest children), and also interact that indicator with the 2010 cohort, to explore whether there were disproportionate changes among the lowest income students. Across both models, the coefficients of interest are the β_4s, which estimate the extent to which

differences in school readiness between cohorts differed by race and SES. These coefficients can be interpreted as changes in school readiness relative to white students and students in the top four SES quintiles, respectively.

Results

Changes in academic skills at school-entry. Figure 1.1 presents the distribution for four sample measures of kindergarten readiness (e.g. reads simple books independently, understands relative quantity). The grey bars show the distribution of these skills in 1998 and the unshaded bars show the distribution in 2010. In all cases we observe the distribution shifting to the right, indicating that in 2010 teachers reported stronger academic skills for their incoming kindergarteners relative to 1998. The pattern is particularly pronounced for letter recognition. In 1998, approximately a quarter of children were rated by their teacher as "not yet" demonstrating that skill, and 15 percent were demonstrated the skill consistently. By 2010 the distribution reversed, with only 15 percent entering not yet recognizing letters and 25 percent doing so consistently. Appendix 1A shows similar figures for all the literacy and math outcomes considered, and shows that to varying degrees this pattern is consistent across all items considered.

In Table 1.1 - Panel A we present changes over the period in our measures of math and literacy proficiency. Model 0 shows results that control only for student age at kindergarten entry and at assessment, and thus addresses our first research question about "raw" changes in school readiness over time. These results mirror the patterns we see in Figure 1.1. Strikingly, students in the 2010 cohort were rated about .25 standard deviations (SD) higher on both math and literacy skills than their

1998 counterparts. Changes were apparent throughout the ability distribution; in both math and literacy we see drops in the percentage of children who were classified as "low proficiency" and increases in the percent that are labeled "high proficiency".

Next we assess to the extent to which these changes are explained by observable covariates. Controlling for demographic characteristics (Model 1), does not explain away the changes over time; rather, differences across cohorts are slightly larger in these models. Counter to our expectations, including measures of preschool participation (Model 2) does not account for *any* of the differences across time for either math or literacy. Adding a rich set of controls for home environment and kindergarten teacher characteristics (Model 3) accounts for roughly 1/5 to 1/3 of the differences across cohorts.

Table 1.1 - Panel B shows changes in teacher-reported behavioral measures over time. Here, we find no overall changes over time on self-control, interpersonal skills or externalizing problem behaviors). However, teachers reported that children in 2010 showed poorer approaches to learning, which measures children's ability to pay attention or adapt to changes in routines. Conversely, they reported that children in the more recent cohort were less likely to demonstrate internalizing problem behaviors, which measure children's shyness or loneliness. Including an increasingly rich set of covariates does not explain any of these differences over time.

Differences in patterns by race and socio-economic status. Figure 1.2 shows how teacher-reported math proficiency changed over time disaggregated by

race and SES. We present the percentage of children that were low proficiency and high proficiency in each cohort. Appendix 1D provides an analogous figure for literacy. A few patterns emerge. First, across both math and literacy, we see that white children are rated as having higher kindergarten readiness skills relative to black and Hispanic children (i.e. more children are classified as "high proficiency" and fewer as "low proficiency"), and similarly that children in the top four quintiles of SES are rated substantially higher than those in the bottom quintile.

The figures also indicate that teachers ranked *all* groups higher in 2010 than in 1998. For all subgroups considered we see drops in the percentage of children at the bottom and increases in the percentage of children at the top. Notably, the overall changes shown in this figure are more pronounced among black children than among other groups. For example, while the percentage of white children who demonstrated low proficiency in math dropped by about 8 percentage points over this period, the same percentage of black children dropped by 15 percentage points, almost double. Conversely, the percentage of white children who were classified as high proficiency in math increased by 9 percentage points from 1998 to 2010, and the corresponding increase for black students was 13 percentage points.

In Table 1.2 we formally examine differences in math and literacy proficiency across subgroups within a regression framework. The top panel explores how changes differ by race/ethnicity. These findings mirror those from Figure 1.2. In particular, while all children experienced increases over this time, the magnitude of changes for black children was disproportionately larger both in literacy and in math, and across each of our summary measures. There is little evidence of

disproportionate changes among Hispanic students, though they did improve on math at a marginally higher rate relative to white peers.

The bottom panel of Table 1.2 shows how changes over time in math and literacy proficiency differ for the poorest children. Again, early math and literacy proficiency increased across the board. We do not find much evidence that children in the lowest socioeconomic quintile showed greater gains over this period, though there is marginal evidence of greater growth in this group for math.

Discussion & Conclusions

This study provides the first nationally representative examination of changes in children's school readiness over time across both academic and behavioral measures. We find that children are arriving at school with a different set of skills than they were in the late nineties, as measured by teacher assessments. In particular, children are entering kindergarten more proficient across a variety of math and reading skills. These changes are sizable. One way to think about the magnitude of the changes in academic readiness observed in the current study is relative to the growth observed from the fall to the spring of kindergarten for the 1998 cohort. In 1998 students' average score on the overall math measure was 2.51 in the fall and 3.62 in the spring. Students in 2010 started kindergarten at an average score of 2.70 in the fall. This suggests that students arriving at kindergarten in 2010 had already learned about 17 percent of what they previously would have learned in kindergarten. For literacy the change is just slightly smaller.

Notably, the changes we observe in the current study were not limited to a single subgroup. We see increases in reading and math readiness across all students

regardless of race or SES, and throughout the distribution of school readiness skills. However, we find particularly large gains in math and literacy proficiency among black children, relative to their white peers. We also find modest but suggestive evidence that math skills increased more among Hispanic children and children in the lowest SES quintile, relative to their white and higher SES peers.

These findings are encouraging and suggest a narrowing in the "school readiness" gap. The patterns are largely consistent with recent research that documents narrowing of race and income-based gaps using *direct measures of cognitive skills* (Reardon & Portilla, 2015). Taken together, these studies suggest that since the late nineties early achievement gaps are narrowing and *simultaneously* that the skills and knowledge children possess when entering school are increasing.

This is an intriguing result and has important implications for the way we structure children's early learning experiences. If children are entering kindergarten with a different set of skills than they once were, it is essential that kindergarten teachers are responsive to these changes. Recent work suggests that there is misalignment between children's skills at school entry and the type of instruction they receive in kindergarten, with children spending too much time on skills they already know (Engel, Claessens, & Finch, 2013). Note that our findings do not imply that kindergarten needs to get more demanding or challenging. Rather, there is a need to better understand what skills children already have at school entry, and how kindergarten can support their development both academically and more broadly.

In the current paper we explored whether increased math and literacy skills at kindergarten entry might have come at a cost in the form of worse behavioral outcomes. Our results here are mixed. We find no differences in teacher-reported behavioral outcomes for self-control, interpersonal skills or externalizing behavior and actually document a *reduction* in internalizing problem behavior. Teachers did, however, rate the 2010 cohort somewhat less favorably with respect to their "approaches to learning," a measure that captures children's eagerness to learn, along with their ability to work independently, persist in completing tasks, and pay attention.

One potential explanation for this shift is the changing nature of kindergarten classrooms. In 2010, kindergarteners spent far more time using textbooks and worksheets, being assessed using standardized tests, and participating in teacher-directed instruction (Bassok, Latham, & Rorem, 2016). They also had far fewer opportunities for child-selected activities, art, music, and hands-on exploration. Perhaps then, kindergarten in 2010 required higher levels of focus, persistence and attention than the kindergarten classrooms in the earlier period. It is not immediately clear what might be driving the drop in internalizing problem behaviors, or why we observe improvements in one behavioral measure but declines in another.

Implications & Next Steps

Our study was motivated by the rapid increase in public investments in early childhood education. We hypothesized that improved access to early childhood learning opportunities has led to improved school readiness, particularly for low-

income and minority children. Although we document trends in school readiness consistent with this hypothesis, our measures of preschool participation failed to explain *any* of the observed changes in school readiness. This surprising result was counter to our expectations.

One potential explanation is that the changes we are observing over time are driven by changes in the *quality* of the child care experiences children have, or changes in the academic focus of these settings. There has been substantial public investment in improving quality both by moving children from informal to formal settings, and by investing in quality enhancements across sectors. Unfortunately, the ECLS-K data do not provide any information about quality measures.

The rich ECLS data allowed us to examine the extent to which a host of covariates measuring child and family demographic characteristics, children's early home environments and kindergarten teacher characteristics explain changes over time in school readiness. When we include all of these covariates we are able to explain about a fifth of the change in math skills and a third of the change in literacy skills. While this is a substantial portion of the change, it raises important questions about what other factors may be driving the increases in school readiness over time. A better understanding of the mechanism driving the observed shifts is essential for understanding how to target new public initiatives or modify existing ones.

More research is also needed to assess whether the changes we observed in teacher-reported outcomes persist as children progress through school, particularly with respect to direct assessments of skills. We cannot fully answer this question until the new ECLS cohort advances through school and a cross-walk is available to

make comparisons of the direct assessments across cohorts. That said, we examined whether differences across our two cohorts were still evident at the end of the kindergarten year (results available upon request). We find that differences across cohorts in teacher-rated early academic skills are *even larger* at the end of kindergarten than they are at the beginning. In addition, the increases observed in the current study are largely mirrored in increasing fourth grade reading and math scores on the National Assessment of Educational Progress (U.S. Department of Education, 2013). These patterns, combined with state reports of rising school readiness over time, bolster our confidence that the teacher-reported measures are capturing "true" changes in children's knowledge at school entry.

References

Allison, P. (2012). Why You Probably Need More Imputations Than You Think | Statistical Horizons. Retrieved February 11, 2016, from http://statisticalhorizons.com/more-imputations

Barnett, W. S., Carolan, M. E., Squires, J. H., & Brown, K. C. (2014). *The State of Preschool 2013: State Preschool Yearbook*. New Brunswick, New Jersey: National Institute for Early Education Research.

Barnett, W. S., Carolan, M. E., Squires, J. H., Clarke Brown, K., & Horowitz, M. (2015). *The State of Preschool 2014: State preschool yearbook*. New Brunswick, New Jersey: National Institute for Early Education Research.

Bassok, D., Finch, J., Lee, R., Reardon, S. F., & Waldfogel, J. (2016). Socioeconomic Gaps in Early Childhood Experiences, 1998 to 2010. *EdPolicyWorks Working Paper*. Retrieved from http://curry.virginia.edu/uploads/resourceLibrary/38_Gaps_In_Early_Experience.pdf

Bassok, D., Latham, S., & Rorem, A. (2016). Is Kindergarten the New First Grade? *AERA Open*, *2*(1), 2332858415616358. http://doi.org/10.1177/2332858415616358

Bassok, D., & Loeb, S. (2015). Early Childhood and the Achievement Gap. In H. F. Ladd & M. Goertz (Eds.), *Handbook of Research in Education Finance and Policy* (2nd ed.).

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & The NICHD Early Child Care Research Network. (2007). Are There Long-Term Effects of Early Child Care? *Child Development*, *78*(2), 681–701. http://doi.org/10.1111/j.1467-8624.2007.01021.x

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, *34*(10), 844.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. *Handbook of Child Psychology*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9780470147658.chpsy0114/full

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *The Teachers College Record*, *112*(3). Retrieved from http://www.tcrecord.org/content.asp?contentid=15440

Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., … Ramey, C. T. (2012). Adult outcomes as a function of an early childhood educational program: an Abecedarian Project follow-up. *Developmental Psychology*, *48*(4), 1033.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, *126*(4), 1593–1660.

Child Trends. (2015). *Early School Readiness: Indicators on Children and Youth*.

Christakis, E. (2016). *The Importance of Being Little: What Preschoolers Really Need from Grownups*. Penguin.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 111–134.

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching Students What They Already Know? The (Mis)Alignment Between Mathematics Instructional Content and Student Knowledge in Kindergarten. *Educational Evaluation and Policy Analysis*, *35*(2), 157–178. http://doi.org/10.3102/0162373712461850

Epstein, A. S., Schweinhart, L. J., DeBruin-Parecki, A., & Robin, K. B. (2004). *Preschool assessment: A guide to developing a balanced approach*. NIEER. Retrieved from http://nieer-www1.rutgers.edu/resources/policybriefs/7.pdf

Fryer, R. G., & Levitt, S. D. (2004). Understanding the Black-White test score gap in the first two years of school. *Review of Economics and Statistics*, *86*(2), 447–464.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system: Manual*. American Guidance Service.

Grimm, K. J., Steele, J. S., Mashburn, A. J., Burchinal, M., & Pianta, R. C. (2010). Early behavioral associations of achievement trajectories. *Developmental Psychology*, *46*(5), 976–983. http://doi.org/10.1037/a0018878

Hecht, S. A., & Greenfield, D. B. (2001). Comparing the predictive validity of first grade teacher ratings and reading-related tests on third grade levels of reading skills in young children exposed to poverty. *School Psychology Review*, *30*(1), 50–69.

Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, *312*(5782), 1900–1902.

Hoge, R. D., & Coladarci, T. (1989). Teacher-Based Judgments of Academic Achievement: A Review of Literature. *Review of Educational Research*, *59*(3), 297–313. http://doi.org/10.3102/00346543059003297

Kilday, C. R., Kinzie, M. B., Mashburn, A. J., & Whittaker, J. V. (2012). Accuracy of Teacher Judgments of Preschoolers' Math Skills. *Journal of Psychoeducational Assessment*, *30*(2), 148–159. http://doi.org/10.1177/0734282911412722

Loeb, S., Bridges, M., Bassok, Fuller, B., & Rumberger, R. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*(1), 52–66.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*(1), 33–51.

Maryland State Department of Education. (2014). *Getting Ready: The 2013-2014 Maryland School Readiness Report*. Baltimore, Maryland.

Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and Classroom Characteristics Associated With Teachers' Ratings of Prekindergartners' Relationships and Behaviors. *Journal of Psychoeducational Assessment*, *24*(4), 367–380. http://doi.org/10.1177/0734282906290594

Minnesota Department of Education. (2013). *Minnesota School Readiness Study: Developmental Assessment at Kindergarten Entrance*.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools and children's life chances* (pp. 91–116). New York, NY: Russell Sage Foundation.

Reardon, S. F., & Portilla, X. A. (2015). *Recent Trends in Socioeconomic and Racial School Readiness Gaps at Kindergarten Entry*. Stanford, California: Center for Education Policy Analysis. Retrieved from http://cepa.stanford.edu/sites/default/files/reardon%20portilla%20school%20readiness%20gap%20trends%2014jan2015.pdf

Schweinhart, L., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry preschool study through age 40*. Ypsilanti: High Scope Press.

Shonkoff, J. P., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.

Stipek, D., Feiler, R., Byler, P., Ryan, R., Milburn, S., & Salmon, J. M. (1998). Good Beginnings: What difference does the program make in preparing young children for school? *Journal of Applied Developmental Psychology*, *19*(1), 41–66. http://doi.org/10.1016/S0193-3973(99)80027-6

Stipek, D., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of Different Instructional Approaches on Young Children's Achievement and Motivation. *Child Development*, *66*(1), 209–223. http://doi.org/10.1111/j.1467-8624.1995.tb00866.x

Teisl, J. T., Mazzocco, M. M. M., & Myers, G. F. (2001). The Utility of Kindergarten Teacher Ratings for Predicting Low Academic Achievement in First Grade. *Journal of Learning Disabilities*, *34*(3), 286–293. http://doi.org/10.1177/002221940103400308

The Build Initiative & Child Trends. (2015). A Catalog and Comparison of Quality Rating and Improvement Systems (QRIS) [Data System]. Retrieved February 10, 2016, from http://qriscompendium.org/

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., Hagedorn, M. C., Daly, P., & Najarian, M. (2013). Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLSK: 2011), User's Manual for the ECLS-K: 2011 Kindergarten Data File and Electronic Codebook (NCES 2013-061). *US Department of Education. Washington, DC: National Center for Education Statistics*.

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., & Najarian, M. (2009). Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K): Combined User's Manual for the ECLS-K Eighth-Grade and K-8 Full Sample Data Files and Electronic Codebooks. NCES 2009-004. *National Center for Education Statistics*. Retrieved from http://eric.ed.gov/?id=ED511826

U.S. Census Bureau. (2015). Table A-1. School Enrollment of the Population 3 Years Old and Over, by Level and Control of School, Race, and Hispanic Origin: October 1955 to 2014. Retrieved April 10, 2011, from http://www.census.gov/hhes/school/data/cps/historical/

U.S. Department of Education. (2013). NAEP - Mathematics and Reading 2013. Retrieved February 12, 2016, from http://www.nationsreportcard.gov/reading_math_2013/#/performance-overview

Virginia Performs. (2015, April 16). Assessing School Readiness in Virginia. Retrieved February 10, 2016, from about:reader?url=http%3A%2F%2Fvaperforms.virginia.gov%2Findicators%2Feducation%2FschoolReadiness.php

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), 2112–2130.

West, J., Denton, K., & Germino-Hausken, E. (2000). America's Kindergartners: Findings from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Fall 1998. Retrieved from http://eric.ed.gov/?id=ED438089

West, J., Denton, K., & Reaney, L. (2001). The kindergarten year. *Washington, DC: US Department of Education*.

Table 1.1 Differences in teacher-rated academic and behavioral outcomes, across cohorts (OLS estimates)

Panel A - Math and literacy proficiency	(0)	(1)	(2)	(3)
Math				
Overall†	0.25***	0.28***	0.28***	0.23***
	(0.02)	(0.02)	(0.02)	(0.02)
Low proficiency	-0.10***	-0.11***	-0.11***	-0.09***
	(0.01)	(0.01)	(0.01)	(0.01)
High proficiency	0.09***	0.10***	0.10***	0.08***
	(0.01)	(0.01)	(0.01)	(0.01)
Literacy				
Overall†	0.23***	0.26***	0.26***	0.18***
	(0.02)	(0.02)	(0.02)	(0.02)
Low proficiency	-0.09***	-0.10***	-0.10***	-0.08***
	(0.01)	(0.01)	(0.01)	(0.01)
High proficiency	0.05***	0.06***	0.06***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)
N	29550	29550	29550	29550
Age	X	X	X	X
Demographics		X	X	X
Preschool variables			X	X
Home environment variables				X
Kindergarten teacher characteristics				X

Panel B - Behavioral outcomes	(0)	(1)	(2)	(3)
Poor self control	0.01*	0.02*	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Poor interpersonal behavior	-0.01	-0.01	-0.01+	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Poor approaches to learning	0.05***	0.05***	0.05***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)
High externalizing behavior	0.00	0.00	-0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
High internalizing behavior	-0.02***	-0.02***	-0.02***	-0.02***
	(0.01)	(0.01)	(0.01)	(0.01)
N	26650	26650	26650	26650
Age	X	X	X	X
Demographics		X	X	X
Preschool variables			X	X
Home environment variables				X
Kindergarten teacher characteristics				X

Note. Each coefficient comes from a separate regression where an outcome was regressed on an indicator for the 2010 cohort. Standard errors are clustered at the teacher level.
†Measure has been standardized to have mean 0 and SD 1.
+p<.1 * p <.05 ** p <.01 ***p <.001

Table 1.2 Differences in teacher-rated math and literacy proficiency across cohorts, by race and SES (OLS estimates)

Panel A. Differences by race

	Math			Literacy		
	Overall†	Low	High	Overall†	Low	High
2010 cohort	0.22***	-0.08***	0.08***	0.21***	-0.08***	0.05***
	(0.03)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)
Black*2010	0.13*	-0.06**	0.04+	0.10*	-0.05*	0.04*
	(0.05)	(0.02)	(0.02)	(0.05)	(0.02)	(0.02)
Hispanic*2010	0.08+	-0.03	0.00	0.04	-0.03	0.00
	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)
Asian*2010	0.08	-0.02	0.01	0.11	-0.01	0.03
	(0.09)	(0.04)	(0.03)	(0.09)	(0.04)	(0.03)
N	29550	29550	29550	29550	29550	29550

Panel B. Differences by SES

	Math			Literacy		
	Overall†	Low	High	Overall†	Low	High
2010 cohort	0.24***	-0.09***	0.09***	0.23***	-0.08***	0.05***
	(0.03)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
SESQ1*2010	0.06+	-0.02	0.00	0.01	-0.03	-0.01
	(0.04)	(0.02)	(0.01)	(0.03)	(0.02)	(0.01)
N	29550	29550	29550	29550	29550	29550

Note. Each coefficient comes from a separate regression where outcomes were regressed on an indicator for the 2010 cohort and interactions between this indicator and either race (omitting white) or the lowest SES quintile (omitting the top four quintiles). Regressions that use race indicators also included "other race" as a category (results not shown). All regressions control for both children's age at kindergarten entry and their age at assessment. Standard errors are clustered at the teacher level.
†Measure has been standardized to have mean 0 and SD 1.
+ p<.1 * p<.05 ** p<.01 ***p<.001

Figure 1.1 Distribution of selected teacher-reported literacy and math skills

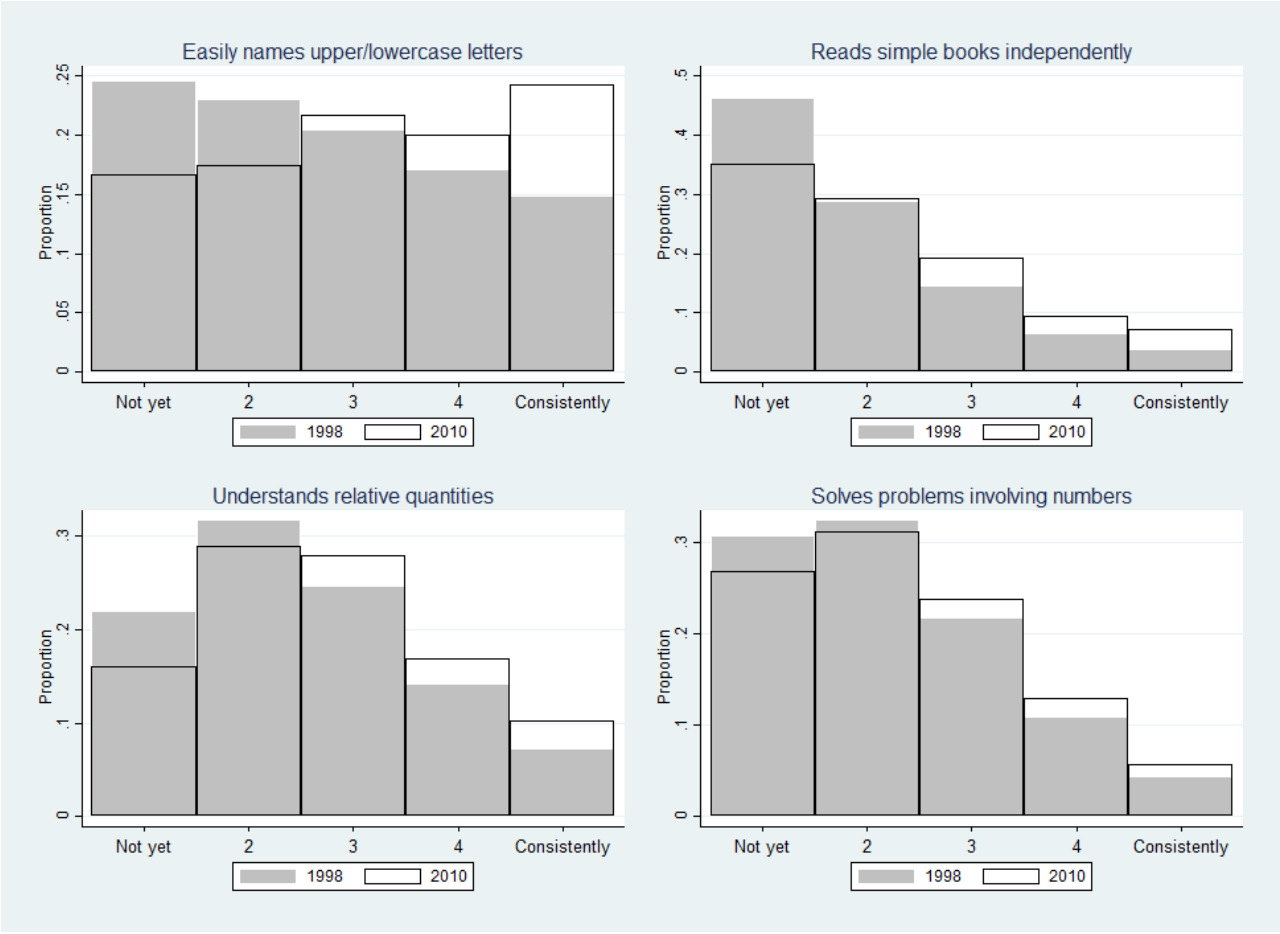
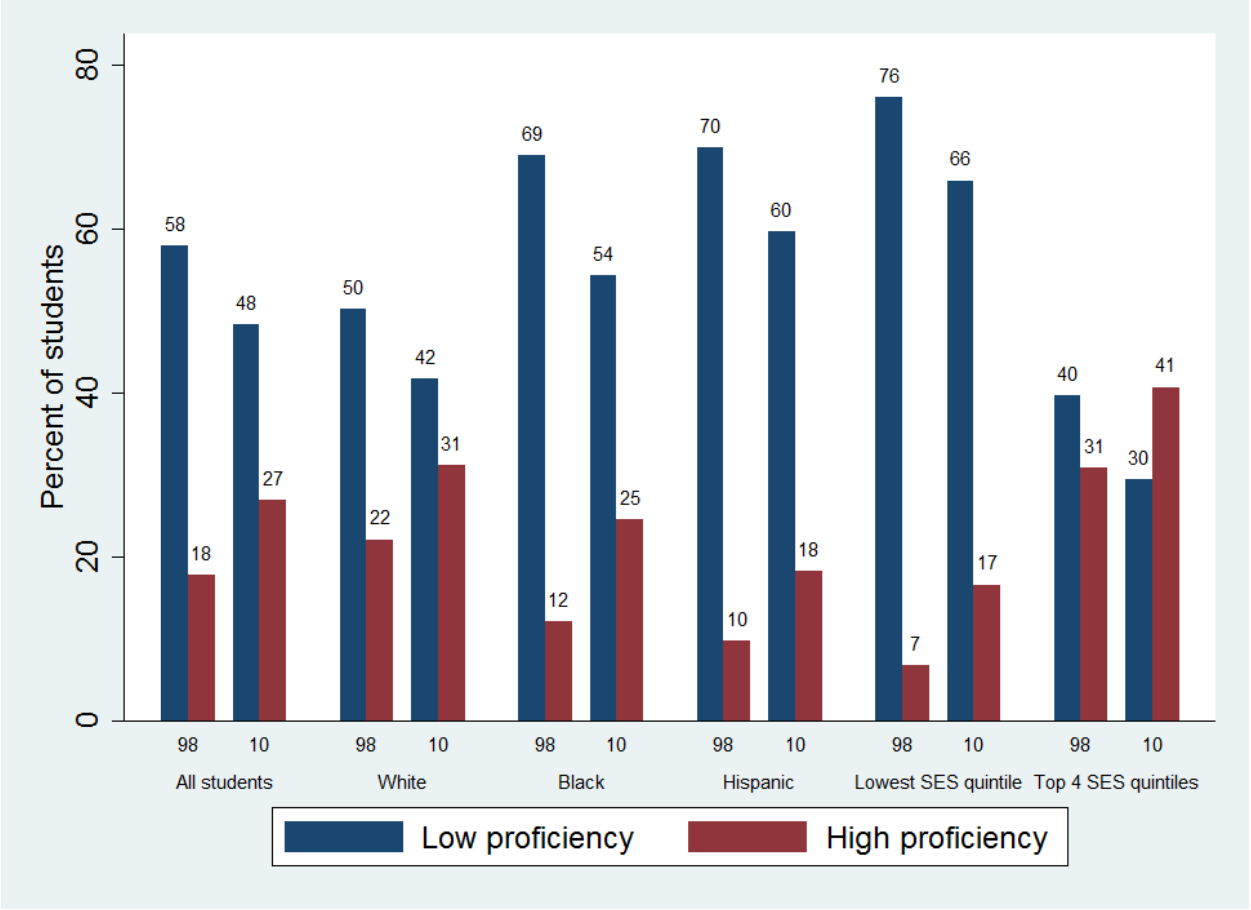


Figure 1.2 Percentage of students who were low and high proficiency in math across cohorts



Appendix 1A. Correlations between teacher-reported measures and direct student assessments

	Direct assessments				
	ECLS-K 2010	ECLS-K 1998			
Math proficiency	Kindergarten	Kindergarten	3rd grade	5th grade	8th grade
Teacher reported					
Overall	0.52	0.49	0.40	0.39	0.37
High proficiency	0.38	0.33	0.26	0.25	0.24
Low proficiency	-0.42	-0.39	-0.34	-0.33	-0.31
Direct math assessment	-	1.00	0.68	0.63	0.58
Literacy proficiency	Kindergarten	Kindergarten	3rd grade	5th grade	8th grade
Teacher reported					
Overall	0.65	0.57	0.46	0.44	0.38
High proficiency	0.48	0.43	0.33	0.31	0.27
Low proficiency	-0.50	-0.43	-0.38	-0.36	-0.31
Direct literacy assessment	-	1.00	0.57	0.52	0.45

Note. Direct assessments were intended to measures broad student ability in math and literacy. These assessments were administered in the fall of the kindergarten year, and in the spring of the 3rd, 5th, and 8th grade years.

Appendix 1B. Covariate descriptive statistics

	1998	2010	
Demographics			
White	0.58	0.52	***
Black	0.16	0.13	***
Hispanic	0.19	0.25	***
Asian	0.03	0.04	***
Male	0.51	0.51	
Age in Aug. of kindergarten (months)	65.76	65.96	***
Speaks language other than English	0.22	0.24	***
Does not speak English	0.03	0.03	
Not U.S. born	0.03	0.03	
Non-citizen	0.02	0.01	
Public school	0.86	0.89	***
Preschool variables			
Attended formal pre-k care+	0.68	0.67	*
Hrs/wk attended pre-k	14.66	15.56	***
Attended publicly funded pre-k	0.28	0.45	***
Attended pre-k/k in same building	0.12	0.17	***
Attended K in school also offering pre-k	0.36	0.50	***

	1998	2010	
Kindergarten teacher characteristics			
Male	0.02	0.02	**
Age	41.66	42.12	**
White	0.91	0.91	
Black	0.07	0.06	
Hispanic	0.07	0.10	***
Asian	0.02	0.02	
Bachelor's degree (no graduate)	0.62	0.53	***
Graduate degree	0.37	0.47	***
Years teaching kindergarten	8.99	8.72	***
Years teaching at current school	9.16	9.13	
Certified in elementary education	0.86	0.86	***
Certified in Early childhood education	0.54	0.54	*
Took coursework in…			
Early childhood education	0.92	0.86	***
Elementary education	0.97	0.94	***
Special education	0.72	0.72	
English as a second language	0.24	0.38	***
Child development	0.97	0.93	***
Methods of teaching reading	0.98	0.95	***
Methods of teaching math	0.95	0.91	***
Methods of teaching science	0.91	0.82	***

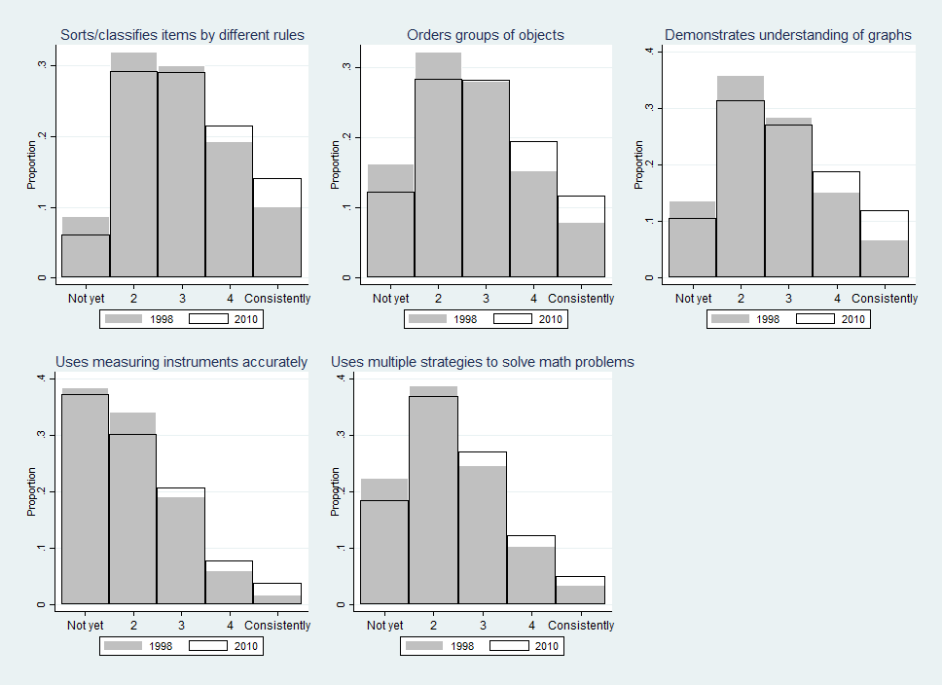
Note. +Head Start or Center-based care. * p<.05 ** p<.01 ***p<.001

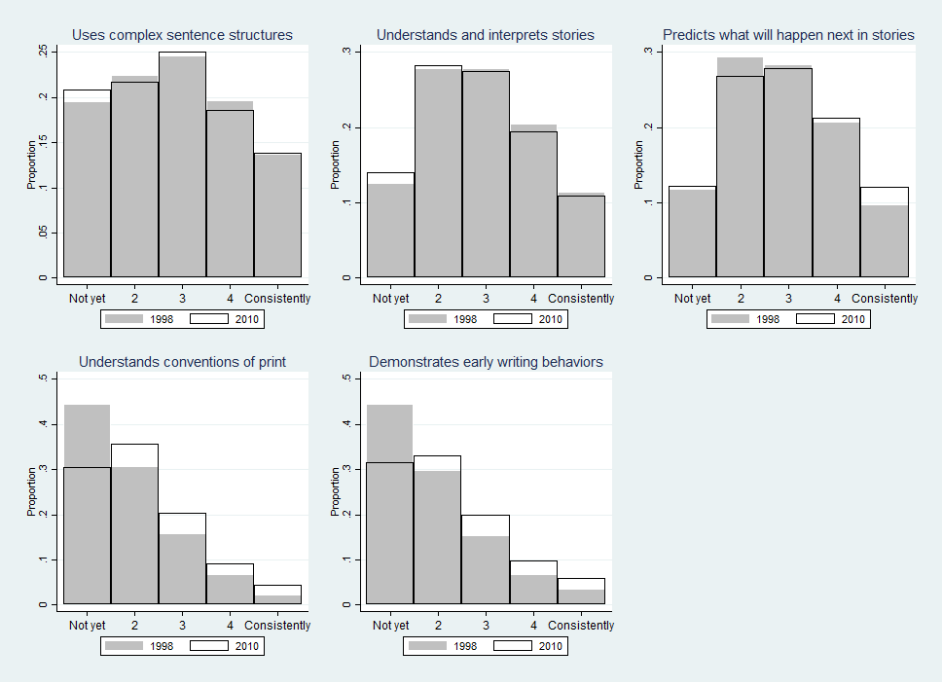
Appendix 1B. Covariate descriptive statistics (continued)

	1998	2010	
Home environment variables			
Proportion of parents rating the following skills "very important" or "essential":			
Knowing most of the letters	0.69	0.82	***
Counting to 20	0.61	0.75	***
Taking turns/sharing	0.95	0.95	*
Using a pencil/paintbrush	0.73	0.83	***
Sitting still/paying attention	0.84	0.86	***
Communicating verbally	0.94	0.96	***
Proportion of parents who report doing the following activities with their children every day:			
Reading books	0.45	0.52	***
Telling stories	0.25	0.40	***
Singing songs	0.45	0.45	
Doing chores	0.53	0.52	
Playing games	0.22	0.24	***
Talking about nature/science	0.10	0.12	***
Building something	0.14	0.17	***
Playing sports/exercising	0.22	0.25	***
Computer use			
Child uses computer at home	0.55	0.74	***
Uses computer every day	0.09	0.11	***
Uses computer for educational pur	0.49	0.64	***
Uses the computer for internet	0.07	0.47	***

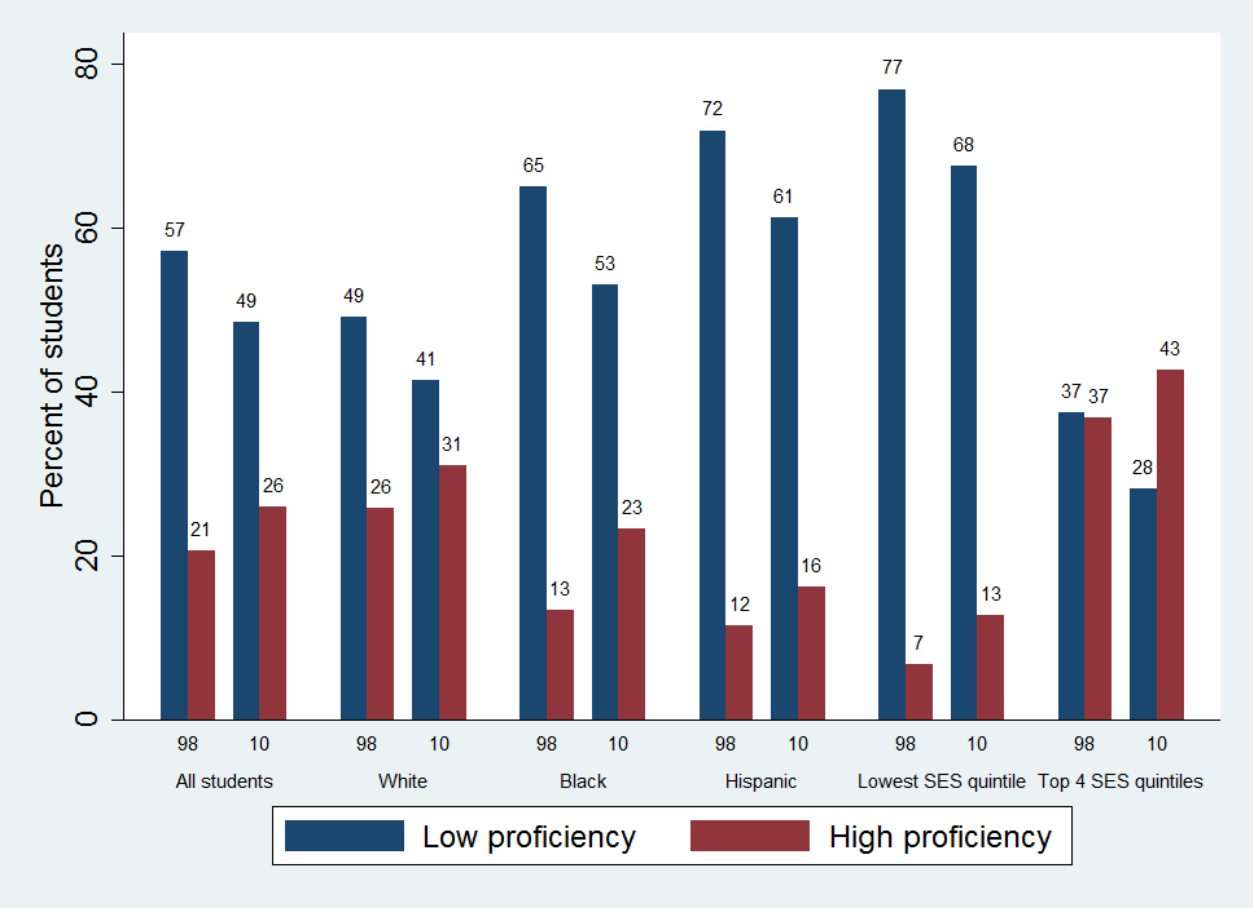
	1998	2010	
Outcome variables			
Academic			
Overall math	2.51	2.70	***
Low math proficiency	0.57	0.50	***
High math proficiency	0.19	0.26	***
Overall literacy	2.46	2.64	***
Low literacy proficiency	0.24	0.19	***
High literacy proficiency	0.22	0.25	***
Behavioral			
Poor self control	0.15	0.17	**
Poor interpersonal behavior	0.17	0.16	
Poor approaches to learning	0.17	0.23	***
High externalizing behavior	0.13	0.13	
High internalizing behavior	0.12	0.09	***

Appendix 1C. Distribution of additional teacher-reported math and literacy skills





Appendix 1D. Percentage of students who were low and high proficiency in literacy across cohorts



Appendix 1E. Differences in teacher-rated student proficiency, across cohorts (bounded OLS estimates)

	(0)	(1)	(2)	(3)
Math				
Overall†	0.14***	0.16***	0.16***	0.11***
	(0.02)	(0.02)	(0.02)	(0.02)
Low proficiency	-0.07***	-0.08***	-0.08***	-0.06***
	(0.01)	(0.01)	(0.01)	(0.01)
High proficiency	0.06***	0.06***	0.06***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)
Literacy				
Overall†	0.24***	0.28***	0.28***	0.19***
	(0.02)	(0.02)	(0.02)	(0.02)
Low proficiency	-0.10***	-0.11***	-0.12***	-0.09***
	(0.01)	(0.01)	(0.01)	(0.01)
High proficiency	0.06***	0.07***	0.07***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)
N	29550	29550	29550	29550
Age	X	X	X	X
Demographics		X	X	X
Preschool variables			X	X
Home environment variables				X
Kindergarten teacher characteristics				X

Note. Each coefficient comes from a separate regression where an outcome was regressed on an indicator for the 2010 cohort. Standard errors are clustered at the teacher level.
†Measure has been standardized to have mean 0 and SD 1.
+p<.1 * p<.05 ** p<.01 ***p<.001

Chapter 2 - The effects of accountability incentives in early childhood education

Daphna Bassok, Thomas Dee, & Scott Latham

CHAPTER 2: ACCOUNTABILITY IN NORTH CAROLINA

Abstract

Quality Rating and Improvement Systems (QRIS), or accountability systems in early childhood, have been rapidly adopted by nearly all states in an effort to improve the quality of early child care. However, we know little about whether two of the primary mechanisms of accountability systems operate as theorized in the early childhood context. Specifically, do differences in QRIS ratings induce either supply or demand side responses in the early child care market? This paper examines these questions in North Carolina using a regression discontinuity design. We find that centers who were quasi-randomly assigned to a lower quality rating responded in multiple ways, both by improving on an observed measure of quality, and by opting out of the rating system entirely. We also find centers that were assigned lower star ratings had lower enrollment five years after ratings were issued, which is consistent with the hypothesis that parents responded to information about quality by selectively enrolling away from centers with lower ratings.

A large and consistent body of research has demonstrated that high-quality early childhood experiences have the potential to narrow achievement gaps and improve children's life trajectories (Bassok & Loeb, 2015; Heckman, 2006; Yoshikawa et al., 2013). Following this evidence, public investments in early childhood education have increased dramatically in recent years. For instance, state spending for preschool increased more than 150% between 2002 and 2015, from 2.4 to 6.2 billion (Barnett et al. 2016). At the same time, some of the best evidence on *scaled-up* early childhood education suggests only modest, short-term benefits (Lipsey et al. 2015, Puma et al. 2012).

Variation in program quality is one of the most common candidate explanations for the at times lackluster impacts of scaled-up public preschool initiatives (Yoshikawa et al, 2013). Despite striking investments in increasing *access* to early childhood education, many of the programs children attend today are low quality, particularly in low-income communities (Bassok, Greenberg, Fitzpatrick, & Loeb, 2016; Bassok & Galdo, 2016). In the face of lower-than-expected returns to increasing *access* to preschool, policymakers have recently focused on improving the *quality* of early child care. For instance, through its Race to the Top – Early Learning Challenge (RTT-ELC), the federal government has competitively allocated roughly one billion dollars to 20 states since 2011, with the explicit goal of improving access to *high-quality* early learning experiences. In 2014, the U.S. Departments of Education and Health and Human Services awarded an additional $250 million to states through their Preschool Development Grants, which required a similar focus on quality.

As part of this wave of support for increased quality in early childhood settings, Quality Rating and Improvement Systems (QRIS) have emerged as a widespread and potentially powerful policy lever. QRIS are accountability systems that aim to improve the quality of the early child care sector. Similar to accountability in other contexts, QRIS programs aim to drive improvements in two ways. First, they establish quality standards and offer financial rewards to programs that meet them. Second, they make information about program quality both public and easily digestible for parents. To this end, they generally use tiered rating systems (e.g. 1-5 stars) to distinguish between settings of varying quality. As of February 2015, 39 states have statewide QRIS systems, and nearly all others are in the planning or piloting phases (QRIS National Learning Network, 2015).

Despite substantial investment in early childhood accountability efforts, there is very little evidence on whether these accountability systems have actually improved quality in early childhood settings, and specifically whether their primary mechanisms work as theorized. This project provides the first such evidence on this high-profile policy initiative by examining North Carolina's Star Rated License (SRL) system, one of the oldest and most well established QRIS in the country. Specifically, we examine the causal effects of receiving a higher versus lower QRIS rating on several observed measures of program quality as well as on future program enrollment. We leverage a regression discontinuity design, taking advantage of the fact that small differences on a continuous measure of program quality led to discontinuous "jumps" in the probability of earning a higher QRIS rating, and address the following research questions:

(1) What is the effect of being quasi-randomly assigned to a lower versus higher QRIS rating on subsequent measures of program performance (e.g. observer ratings, staff education/experience)?

(2) What is the effect of a lower vs. higher QRIS rating on programs' subsequent enrollment and likelihood of closure?

(3) Do these effects vary across different types of programs?

In addressing these questions we provide the first credibly causal evidence on two key mechanisms underlying the QRIS theory of change. We find evidence of effects of these ratings on both future program quality and enrollment. In particular, we find that some providers who received a lower rating responded by improving their scores on an observational measure of quality, while others opted out of the rating process entirely. We find that both of these effects are concentrated among centers with enrollment below the median. We also find that centers who received lower ratings had significantly lower enrollment five years after ratings were issued, which is consistent with parents responding to ratings by selectively enrolling away from lower rated centers.

Background

The importance of high quality early child care

For investments in early childhood education to yield long term benefits, programs must offer children high-quality, engaging environments. The studies most commonly cited to demonstrate long-lasting effects from early childhood interventions involved intensive, high-quality programs (Schweinhart et al., 2005; Campbell et al., 2012). While not as intensive, the present-day, scaled-up preschool programs that have shown promise, such as Oklahoma's Universal Preschool

Program or Boston's Preschool Program, also emphasize the importance of high quality (Yoshikawa et al., 2013).

Measuring quality

Measuring the "quality" of early child care providers is difficult given the multiple goals of child care. Unlike the K-12 context, where quality is often defined based on the "value-added" schools add with respect to children's learning outcomes, in early childhood settings quality measures are generally divided into "structural" and "process" quality. Structural quality includes features that are straightforward to measure and regulate, such as teacher education and experience levels, class size, or staff-child ratios, and that are hypothesized to facilitate high quality learning experiences for young children. Process measures aim to more directly capture the quality of the child's experience in the classroom (e.g. the sensitivity of the teacher in interacting with children, or the extent to which the classroom is stimulating, engaging, and positive). Measures of process quality are generally administered during a classroom visit. A large body of research has demonstrated that measures of process quality in particular are related to children's developmental outcomes in preschool (Araujo, Carneiro, Cruz-Aguayo, & Schady, 2014; Hamre & Pianta, 2005; Howes et al., 2008; Mashburn et al, 2008; Sabol, Hong, Pianta, & Burchinal, 2013).

Despite the importance of quality, and the dramatic increase in *access* to publicly-funded early child care in recent years, a large portion of existing child care settings and preschool programs have been characterized as low-quality or inadequate (Helburn, 1995; Wrigley & Dreby, 2005; Bassok et al., 2016). For instance, a study of pre-kindergarten and early education across 11 states found

that classroom quality, and instructional quality in particular, is generally low, and that children have relatively few meaningful interactions with adults and other children throughout the day (Early et al., 2005).

QRIS as a tool for improving quality

Traditionally, early child care quality has been regulated by establishing "floors" for quality, or minimum requirements that care providers must meet. QRIS are attempts to incentivize programs to move above these floors. Typically, QRIS establish multiple "tiers" of quality (e.g. 1-5 stars) with clear benchmarks for each. These benchmarks generally combine measures of structural and process quality. Most state QRIS also include an observational component in which providers are assessed during a classroom visit. Providers are then assigned quality ratings based on the benchmarks that they meet.

These ratings are theorized to drive improvements in quality in two ways. First, providers receive direct financial incentives for meeting higher quality benchmarks (e.g. subsidy reimbursement rates; merit awards). At the same time, states and/or local organizations generally provide support for quality improvement such as professional development or technical assistance (QRIS National Learning Network 2015).

Second, QRIS collect and disseminate simplified information about child care quality to parents and other stakeholders, who are often unable to discern quality on their own (Cryer & Burchinal, 1997; Mocan, 2007). It is theorized that parents will then use this information to "vote with their feet," selectively enrolling away

from lower quality settings, and putting additional pressure on care providers to improve quality or face decreased enrollment or closure.

There is scant evidence as to whether either of these mechanisms is effective in the early childhood context. Instead, most of the existing research on QRIS has focused on establishing the validity of QRIS ratings by comparing them to other measures of quality or to child outcomes (Sabol et al., 2013; Sabol & Pianta, 2014). One small experimental study of a local QRIS in Washington state provides some evidence that coaching and improvement as part of a QRIS improved observed child care quality over a 6 month period (Boller et al., 2015). However, the study did not provide any information about whether care settings are responsive to financial incentives, or whether parents are sensitive to differences in quality ratings. Further, no studies have examined whether QRIS can meaningfully alter measures of program performance or enrollment when fully implemented at scale.

To answer these questions, it is essential to separate parent/provider responses that are driven by differences in quality ratings from those that are driven by actual differences in quality. For example, if providers that receive lower ratings experience a subsequent drop in enrollment, it could be either because parents are making decisions based on ratings or because they are acting on independent information about the relative quality of providers. To get around this endogeneity issue, an ideal experiment might randomly assign providers *of the same quality* to receive either a high or low quality rating. We could then be confident that subsequent differences in either program quality or enrollment were driven by differential responses to these different ratings. As we describe in detail below, this

is precisely the experiment that we approximate with the current regression discontinuity design.

To provide context for how accountability systems may function in early childhood, we now turn to findings from two related areas of research: the K-12 literature on accountability systems and their impact on student achievement and the literature on information and school choice.

Accountability systems and program improvement

There is a fairly substantial body of evidence that K-12 accountability systems such as the federal No Child Left Behind can yield meaningful program improvements as evidenced by gains in student achievement (Dee & Jacob, 2011; Wong, Cook, & Steiner, 2015). For example, a 2011 report from the National Research Council concluded that school-level incentives like those in NCLB raised achievement by about 0.08 standard deviations (particularly in elementary-grade mathematics). At the same time, design elements of accountability systems have important implications. For instance, Hanushek & Raymond (2005) found that accountability systems based on information alone, rather than information coupled with incentives, did not influence the performance of K-12 schools. There are also concerns that when incentives created by accountability systems are poorly designed and targeted, they may have neutral, negative, or unintended consequences (National Research Council, 2011).

Different people and organizations also respond differently to the same incentives. In Washington, D.C. for instance, teachers responded to a threat of dismissal both by improving performance and by exiting the district (Dee &

Wyckoff, 2015). Dee & Jacob (2006) also found heterogeneous responses to high school exit exams, such that some students were incentivized to perform better and others were more likely to drop out. This suggests that accountability systems are likely to have diverse effects, especially when applied to a diverse set of providers. Nevertheless, Figlio & Loeb's (2011) review suggests that the majority of research finds positive effects of accountability efforts on student achievement. Taken together then, the K-12 accountability literature suggests that the movement to bring accountability systems to early childhood settings should be viewed with cautious optimism.

At the same time, there are reasons why accountability in the early childhood context may differ from the K-12 case. First, the early childhood landscape is much more diverse and fragmented than the K-12 sector in terms of goals (e.g. work support for parents, developmental supports for children), sector (e.g. school-based pre-kindergarten, federally-funded Head Start, licensed private child care settings, and family day care homes), cost (e.g. free, subsidized, or unsubsidized) and intensity (half- versus full-day, 9-month versus 12-month). Second, given the difficulty of reliably and cost-effectively assessing children early in childhood, there are serious reservations about "outcomes based" accountability in early childhood, and such a system would be quite difficult to implement (Snow & Van Hemel, 2008). In light of the heightened policy interest and investment in early childhood accountability systems, there is need for more research specifically exploring the impacts of accountability systems within an early childhood context.

Information in the market for schools

One way that accountability systems aim to improve program quality is by providing parents with clear information about quality, so that they can be informed consumers. Parents generally struggle to differentiate between higher- and lower-quality preschool programs, and when asked to assess the quality of their own child's care arrangement they tend to rate nearly all aspects very highly (Cryer & Burchinal, 1997; Peyton, Jacobs, O'Brien & Roy, 2001; Van Horn, Ramey, Mulvihill, & Newell, 2001). For instance, Raikes, Torquati, Wang & Shjegstad (2012) find that 29 percent of parents give their child's provider an overall rating of "perfect" and another 45 percent rate the provider "excellent." This pattern is clearly at odds with research findings about the highly variable quality of care available. A handful of existing studies have compared parents' ratings of care quality to observer ratings of the same measures (Cryer & Burchinal, 1997; Cryer, Tietze, & Wessels, 2002; Mocan, 2007). Two key themes from this research are that parents substantially overestimate the quality of their child's care arrangement relative to ratings by outside observers, and that correlations between parent and observer ratings are modest. Mocan (2007) argues that parents, particularly single parents or parents with lower levels of education, are unsuccessful in evaluating the true quality of ECE arrangements and suggests that providing information about program ratings could be a potential policy solution for addressing this "information asymmetry."

A compelling research base suggests that parents are responsive to clear information about K-12 school quality (Friesen, Javdani, Smith & Woodcock, 2012; Koning & van der Wiel, 2013). For instance, Hastings & Weinstein (2008) provide

experimental evidence that parents who received simplified information about school quality selected higher-quality schools for their children, and that these choices in turn led to improvements in children's test scores. A study in Texas found that parent decisions to exit charter schools were significantly related to the schools' performance on state accountability tests (Hanushek, Kain, Rivkin, & Branch, 2007). The provision of information about quality has also driven changes in other contexts. For example, Pope (2009) found that hospitals with higher U.S. News and World Report rankings attracted significantly more patients. Other work has found that publicized "report cards" for health care providers have improved the match between patients and physicians (Dranove et al., 2002) and significantly changed consumer health care decisions (Dafny & Dranove, 2008; Jin & Sorensen, 2006). Finally, Jin & Leslie (2003) found that publicly displayed hygiene grade cards in restaurants caused consumers to become sensitive to hygiene quality, leading to increased health inspection scores and reductions in food borne illness. In general, the evidence strongly suggests that consumers are responsive to information about quality. Given the theoretical promise of accountability programs in early childhood and of providing information about quality to parents, we now turn to our empirical evaluation in North Carolina.

The North Carolina Context

Advantageous features of North Carolina's QRIS

North Carolina provides a compelling context to study the effects of a large-scale early childhood accountability effort for several reasons. First, North Carolina's Star Rated License (SRL) program is one of the oldest and most established QRIS in the

country. It was instituted in 1999, and has been operating in its current form since 2005. North Carolina spends more than $13 million yearly to administer its QRIS, more than any other state, and maintains nearly a decade of rich, program-level data on star ratings as well as the underlying quality measures that go into calculating the ratings. The program has all the key features of a mature QRIS including (1) well-defined quality standards linked to financial incentives; (2) support for program improvement through technical assistance and local partnerships; (3) regular quality monitoring and accountability and; (4) easily accessible quality information provided to parents (Tout, Zaslow, Halle, & Forry, 2009; Zellman & Perlman, 2008; The Build Initiative and Child Trends, 2015).

Second, in contrast to most state-run QRIS, the vast majority of licensed child care centers participate in the SRL program, including all Head Start centers, all state pre-kindergarten programs, and most centers that operate in local public schools. Whereas most state-run QRIS require centers to "opt in," in North Carolina, all non-religious centers are automatically enrolled at the lowest (i.e. one star) level when they become licensed. Settings may then apply for higher ratings after a temporary waiting period. In total, roughly 88% of licensed centers received star ratings in any given year. The 12% that do not receive star ratings consist primarily of religious sponsored facilities (10%), with a smaller number having temporary/provisional licenses (2%). This high rate of participation is crucial for understanding how QRIS function when applied to a diverse universe of programs, rather than a small and self-selected portion of the market.

A third and particularly crucial feature of North Carolina's rating system is that programs' star ratings (which range from 1-5 stars) are determined in part by a continuous measure of observational classroom quality. In contrast to other components of the QRIS, which are scored as discrete measures, this continuous measure of quality allows us to leverage a regression discontinuity design. Specifically, providers must exceed a set of thresholds on the observation metric to attain credit toward a higher star rating. This means that small differences in programs' observation scores can make the difference between earning a higher or lower star rating (e.g. "3 star" versus "4 star"). We leverage these contrasts to estimate the causal impact of receiving a higher vs. lower star rating on subsequent measures of program quality and on enrollment. Taken together, the North Carolina context and data provide a fertile setting to conduct the first study on the effects of a scaled-up early childhood accountability system.

The Star Rated License system

In this section, we provide an overview of how quality ratings are determined in North Carolina, focusing on the specific features of NC's QRIS that are crucial for understanding and interpreting this research. A complete description of how quality ratings are determined is outside of the scope of this paper, but interested readers can learn about this in detail at North Carolina's Division of Child Development and Early Education website (ncchildcare.nc.gov).

In North Carolina, center-based care providers are rated on a scale from 1-5 stars. New providers are automatically enrolled at the lowest level (i.e. 1 star) upon becoming licensed (religious centers may opt not to receive a star rating). After 6

months, centers may apply to be evaluated for a higher star rated license (i.e. 2 to 5 stars). The number of stars that are awarded is based on an underlying 15 point scale as follows:

Total number of points:	Star rating
0-3	1 star
4-6	2 stars
7-9	3 stars
10-12	4 stars
13-15	5 stars

These 15 possible points are awarded in integer increments (i.e. no partial points) and are primarily earned across two subscales, each worth up to 7 points. The first, "program standards" (≤ 7 points), includes structural measures of quality such as administrative, personnel, and operational policies, along with staff-child ratios and square footage requirements. Each of these measures is scored on a discrete scale. As described in detail below, program standards also includes an observational component scored on a continuous scale that combines measures of structural and process quality.

The second subscale, "staff education" (≤ 7 points), is determined by the education and experience levels of administrators, lead teachers, and the overall teaching staff. For instance, centers receive higher scores for a staff with more years of early childhood teaching experience or more advanced training in the field. Each center can also receive an additional "quality point" by meeting at least one of a variety of other education or programmatic criteria (e.g., using a developmentally appropriate curriculum, combined staff turnover of ≤ 20%, 75% of teachers/lead

teachers have minimum 10 years early childhood experience). Each component of the staff education subscale is scored on a discrete scale.

Environmental Rating Scales. A particularly important feature of the SRL system for the purposes of this study is that providers are eligible for more points in the "program standards" subscale, and in turn higher star ratings, depending on how they score on a measure of quality called an Environment Rating Scale (ERS).[4] ERS are widely-used observation tools currently included in 30 state-run QRIS throughout the country. They are designed to be broad measures of classroom quality, and to incorporate both structural features of the classroom (e.g. space and layout, daily schedules) as well as measures of "process" quality like student-teacher interactions and classroom activities. During the rating process, assessors conduct site visits where they randomly select a third of classrooms to be rated, including at least one classroom for every age group served (i.e. infants/toddlers, 3-4 year olds, school aged children). Assessors, who typically have a background in early childhood or a similar field, spend a minimum of 3 hours in each classroom, recording notes on a wide variety of interactions, activities, and materials. They also spend 30-45 minutes interviewing the lead classroom teacher. In contrast to other components of the QRIS, which are scored on discrete scales, ERS scores are scored

[4] In North Carolina, the Division of Child Development contracts with the North Carolina Rated License Assessment Project (NCRLAP) to conduct these assessments. Centers must submit a request to be rated, and they receive a four week scheduling window during which assessors may visit at any time. NCRLAP stresses the importance of evaluations occurring on a "typical day," and to this end centers are allowed to designate up to five days as non-typical days during which assessments will not occur. ERS ratings remain valid toward a center's star rating for 3 years, and North Carolina provides one free assessment every three years. Centers wishing to re-rated sooner must cover the cost of assessment on their own (North Carolina Rated License Assessment Project, n.d.).

using a continuous scale from 1-7, with 1 indicating "inadequate" quality, and 3, 5, 7 indicating "minimal," "good," and "excellent" quality respectively (The Build Initiative & Child Trends, 2015).[5] In North Carolina, these ratings are valid for three years.

Centers are not required to receive ERS ratings, but those that elect to be rated are eligible for higher overall star ratings. For example, centers who opt to forego an ERS rating can earn a maximum of just 2 points in "program standards" (out of 7), meaning the highest total score they could receive is 10 points (assuming they earn *all 7* "education standards" points as well as an additional quality point). This means a center choosing not to receive an ERS rating is automatically disqualified from receiving a 5 star rating (which requires 13 points), and must earn every other point possible to receive even a 4 star rating (which requires 10 points). In practice, most centers opt to receive ERS ratings, and the percentage has been increasing over time, from 52% in 2008 to 66% by 2014.

Among centers that elect to receive an ERS rating, both the *average* score that a center receives across classrooms and the *lowest* score received can influence the total number of points earned. Centers earn additional points by exceeding a series of thresholds along each of these. For instance, a center with a lowest ERS

[5] Four different versions of the ERS are used depending on the age of children and the type of care setting. In particular, care settings may be rated using the Early Childhood Environment Rating Scale - revised (ECERS-R, 47 items; Harms, Clifford & Cryer, 1998), the Infant/Toddler Environment Rating Scale - revised (ITERS-R, 39 items; Harms, Cryer, & Clifford, 2003), the School-Aged Care Environment Rating Scale (SACERS, 49 items; Harms, Jacobs, & White, 1996), or the Family Child Care Environment Rating Scale - revised (FCCERS-R, 38 items; Harms, Cryer, & Clifford, 2007). Although the scales are tailored to specific age groups, each is scored on the same 1-7 scale, and contains measures of basic care provision, physical environment, curriculum, interactions, schedule/program structure, and parent/staff education.

score above 4.0 can earn a maximum of 6 points on program standards, while a center with a lowest score below 4.0 can only earn a maximum of 2 points. Similarly, a center with an average ERS score of 4.5 is eligible for up to 4 points on program standards, whereas a center that receives just below a 4.5 is only eligible for 3 points (see Appendix 2A for full details of how program standards scores are calculated). This means that small, and arguably random, differences in ERS ratings can be the difference between a provider earning a higher or lower point total on the program standards scale. Since each point constitutes roughly 1/3 of a star, these same small differences can lead to meaningful differences in the probability of earning a higher vs. lower star rating.

Differences between higher and lower star ratings. ERS ratings are critical to the current study because they provide plausibly exogenous variation in star ratings. In turn, star ratings are hypothesized to be critical components in the QRIS theory of action. First, providers receive direct financial incentives for meeting higher levels of quality. In particular, they receive higher per-student reimbursements for subsidy-eligible children for every additional star they earn. These increases vary by county and by the age of children served, but in most cases they are substantial. For instance, in 2007, a 5 star center received an 11% higher reimbursement on average per subsidy-eligible student than a 4 star center. A 4 star center received a 5% higher reimbursement than a 3 star center, and strikingly, a 3 star center received a 35% higher per-student reimbursement than a 2 star center. These differences in subsidy rates are hypothesized to incentivize lower-rated

centers, particularly those who enroll many subsidy-eligible children, to improve their quality in order to qualify for higher rates.

Next, star ratings are publicly available, and may influence parents' choices about where to enroll their children. Since inception, North Carolina has implemented multiple strategies to increase awareness of the Star Rated License program. This includes publishing all star ratings through a searchable tool on North Carolina's Department of Health and Human Services website, distributing posters, business cards and postcards with the web address for this tool, arranging for media coverage of highly rated centers, and requiring star rated licenses to be displayed prominently within each center (National Center of Child Care Quality Improvement, 2015; See Appendix 2B for an example of a star rated license). Many centers also post star ratings on Facebook.

Given the concrete incentives that care providers face, and the efforts in North Carolina to publicize quality information, this context provides an opportunity to evaluate two important theorized mechanisms of accountability in early childhood in a fully-implemented, model QRIS. We hypothesize that centers who receive lower ratings will likely focus on making improvements in their ERS ratings, since small improvements along this dimension are likely to lead to higher star ratings. Since ERS ratings are valid for three years, we first expect to see improvements along this measure a minimum of three years after the initial rating. We also hypothesize that lower rated centers will face a drop in enrollment as a result of lower demand, though this will depend both on whether parents are aware of star ratings and whether they use them to make decisions about child care (both

of which are unobserved). Importantly, although we expect that decreases in enrollment are likely to be driven by parental decisions, our research design does not allow us to untangle whether any changes are due to demand side responses from parents or supply side responses from providers.

Responses to lower star ratings are likely to vary across centers. In some cases, centers may respond by improving their quality, while in others, centers may elect to do nothing, or even disengage from the rating process. We may also observe differences between centers with lower and higher enrollment, as it may be less costly to improve quality among smaller providers.

Method

Data

This analysis leverages setting-by-year data for all licensed child care centers in the state of North Carolina in the years 2007-2014. These data, drawn from the NC Department of Health and Human Services, span nearly the entire period since the last major revision to North Carolina's rating system in 2005, and include all licensed child care centers in North Carolina (N=6,929 unique settings across the entire panel). For each center-year observation, we have data on county of operation, auspice (independent center, Head Start, etc.), enrollment, and capacity. We also have detailed information about program quality as measured through the SRL program, including overall star ratings, program standards and staff education scores, and whether each setting earned a quality point. The data also include the underlying components that comprise the points, including ERS scores, various

structural features of the classroom (e.g. child-staff ratio, space requirements), and indicators of staff education/experience levels.

Constructing the intent-to-treat (ITT) sample. Our design leverages the relationship between ERS scores and star ratings to estimate the causal impacts of receiving higher versus lower star ratings. This is somewhat complicated by the fact that North Carolina revised its QRIS regime in 2005, and that the rollout was staggered across multiple years. In particular, ratings that took place on or after January 1, 2006 were scored under the new regime, but pre-existing centers had until January 1, 2008 to transition to the new system (NC Division of Child Development and Early Education, n.d.). Our data begin in 2007, and because ratings are valid for multiple years, some of the ERS ratings we observe potentially reflect ratings from the previous regime. It is important to exclude all such ratings from our analysis, as they were scored under a different system and do not reflect the same treatment contrast (i.e. ERS ratings were not related to star ratings in the same way).

To address this issue, we define our ITT sample using a center's first rating under the new rating regime. Specifically, we rely on recorded ERS visit dates where possible (about 47% of observations), and classify all recorded visits that occurred in 2007 or later as belonging to the new regime. In cases where ERS visit dates are not recorded, we use several decision rules to determine which ERS ratings were scored under the new regime. Because ERS ratings are valid for three years, we assume that ratings were initially conducted in 2007 if we observe the same rating throughout the years 2007-2009. In cases where we observe a rating in 2008 or

2009 that differs from the 2007 rating, we include the first *changed* rating in our ITT sample.

Our analysis tracks annual center outcomes for as many as five years after the 2007-2009 baseline period. Our data include 6,929 unique centers but a portion of these centers (N = 1,063) were not observed in our baseline period and are, therefore, excluded from our RD analysis. We also excluded 844 centers that never had a star rating (i.e. operated under a religious sponsored, temporary, or provisional license). In addition, we exclude 1,865 centers that had a star rated license but did not opt to receive an ERS rating. This leaves us with a final sample of 3,157 centers in our ITT sample.

Table 2.1 presents descriptive statistics for this sample in the baseline year (T) and for subsequent years through T+5. At baseline, 45% of centers were operated independently, a little more than a quarter were in local public schools, and 9% were Head Start centers. The vast majority (97%) had earned at least a 3 star rating, and 44% had earned a five star rating. The average ERS rating was 5.21, indicating relatively high quality on average across the sample.

It is important to note that because we exclude both centers with no star ratings (primarily religious-sponsored) and centers who chose not to receive an ERS rating, we are tracking only 56 percent of the licensed centers operating in North Carolina during our baseline period. This sample restriction has implications for the external validity of our findings, as the omitted providers differ considerably from those in our sample. Specifically, Appendix 2C shows how centers from our sample compare to omitted providers in the years 2007-2009. For instance, in 2007 53% of

excluded centers were independently operated, compared with 44% of centers in the sample. As expected, excluded centers were more likely to be religious sponsored (21%, compared with 8%), and less likely to be located in local public schools (17% compared with 27%). Only 1% of excluded centers were Head Start centers, compared with 10% of centers in the sample. Excluded centers also had lower average enrollment both overall and relative to capacity. As to be expected, these centers also had *much* lower star ratings on average. In general, centers that are in the sample are likely to be of higher average quality than those that are excluded. However, even this sub-sample of QRIS participants represents a larger portion of the state's centers than the portion included in many state's QRIS (The Build Initiative & Child Trends, 2015).

Identification strategy

Our RD analysis compares outcomes among programs whose average ERS rating at baseline is above or below an ERS cutoff. This contrast implies a fuzzy regression discontinuity design, as programs that are just below this cutoff (i.e. "treated" centers) are significantly less likely to receive a higher star rating compared to programs just above the cutoff (i.e. "untreated" centers). In this design, treated centers who receive lower star ratings based on being below the RD threshold may face incentives to improve quality both directly through higher subsidy rates and indirectly through parent enrollment decisions. As is common practice (Lee & Lemieux 2010, Schochet et al. 2010), we employ a combination of graphical and statistical evidence in our analysis. We estimate the magnitude and statistical

significance of receiving a higher vs. lower star rating using least-squares specifications that take the following form for outcome Y_i associated with center i:

$$Y_i = \gamma \mathbb{1}(S_i \leq 0) + k(S_i) + \alpha_i + \varepsilon_i$$

The variable S_i is the assignment variable (i.e. the center's average ERS rating) centered around 4.5, the focal RD threshold in the current analysis[6], and k is a flexible function of the centered assignment variable. We condition on a fixed effect, $\alpha_{i,}$ for the specific year in which a center's ERS rating occurred (2007-2009), and ε_i is a mean-zero random error term. We report robust standard errors throughout. The parameter of interest, γ, identifies the decreased likelihood of receiving a higher star rating for centers just below the 4.5 threshold compared to those just above it.

To examine effects on program quality, we consider differences in future star ratings, ERS scores, and other indicators of quality measured as part of NC's QRIS such as staff-child ratios and space requirements. We also consider enrollment (both total and as a proportion of capacity), and likelihood of closure. As a measure of whether centers responded by disengaging with the rating process, we consider the likelihood that centers were still open but no longer had a valid ERS rating. Finally, we examine the heterogeneity of these effects both by enrollment (high vs. low enrollment) and by auspice (Head Start centers/local public schools vs. other types of centers). We report intent to treat (ITT) results throughout.

[6] The SRL system also implies other candidate thresholds that may be leveraged using a regression discontinuity. Specifically, centers are eligible for more QRIS points when their *lowest* ERS rating across classrooms exceeds either 4.0 or 5.0, or when their *average* ERS rating exceeds 4.75 or 5.0. We ultimately focus on the average ERS rating as a forcing variable to address the potential manipulation concerns discussed below. We focus on the 4.5 cut-off primarily because it offers the strongest "first stage" relationship (i.e. this cutoff is most strongly related to star ratings).

Assignment to treatment. A regression discontinuity design relies on the fact that small changes in an assignment variable lead to large and discontinuous changes in treatment status. In the North Carolina context, this means that small differences in ERS ratings must lead to discontinuous probabilities in earning a higher star rating. For this project, we leverage the fact that earning an *average* ERS score just below 4.5 makes a center significantly less likely to earn a higher star rating. In Figure 2.1, we illustrate two "first-stage" relationships implied by the 4.5 threshold. Here, we organize centers into bins of size .2 on either side of the threshold, and show the proportion of centers who earned a 3+ or 4+ star rating in each bin. We superimpose regression lines from parametric estimates with quadratic splines.

Figure 2.1 shows that in North Carolina, centers whose average ERS rating was ≤ 4.5 were 12 percentage points less likely to receive a 3+ star rating than those just above 4.5. These centers were also 28 percentage points less likely to receive a 4+ star rating. In Table 2.2 we also present "local linear" first stage estimates, including linear splines for the full sample and for increasingly narrow bandwidths down to the recommended Imbens & Kalyanaraman (2009) bandwidth of 1. These estimates are quite similar to the quadratic specification. Note that these patterns reflect a "fuzzy" rather than "sharp" regression discontinuity, in that the probability of treatment jumps discontinuously, but does not jump from 0 to 1. The importance of this distinction is that the resulting estimates may generalize only to "compliers," that is, centers for whom an additional point resulted in an increased star rating.

Potential for manipulation of the assignment variable. A key identifying assumption of regression discontinuity designs is that actors are not able to precisely influence their score on the assignment variable relative to the RD threshold. In this context, either care providers or raters could be a source of such manipulation. For instance, if some providers could precisely determine the ERS ratings they would receive prior to being rated, and if ERS ratings are relatively easy to improve, then these providers could essentially choose where they fall relative to the ERS cutoffs. In practice, this is unlikely to occur. Although providers are able and encouraged to conduct self-assessments on the ERS survey, these self-assessments are not accurate enough to provide precise information about the ERS ratings providers will ultimately receive.

The second source of potential manipulation comes from raters themselves. This could happen if raters choose to "bump up" ERS ratings for some providers that fall just below an ERS threshold. This issue is of greater concern, because raters are certain to have knowledge of both the scores they are assigning and of the implications of scoring above or below particular thresholds. This concern is primarily what drives our decision to focus on the "average," rather than the "lowest," ERS rating as an assignment variable. Specifically, since multiple ERS ratings are averaged together, no single rater can exert precise control over the final rating.

Conceptually, then, precise manipulation of the assignment variable is unlikely in this context. To corroborate this empirically, we examine a standard battery of tests for manipulation. First, we perform a visual inspection of the density

of the assignment variable. Here we construct binned density plots, organizing the assignment variable into .05 and .025 point bins on either side of the 4.5 threshold (Figure 2.2 – Panel 1). These plots suggest no discontinuity in density at the 4.5 threshold. We test for a discontinuity formally using the commonly employed McCrary density test (McCrary 2008, Figure 2.2 – Panel 2) as well as a newly developed procedure that improves upon the McCrary test in some ways (Cattaneo, Jansson, & Ma, 2015 – results not shown).[7] With both tests, we fail to reject the null hypothesis of no discontinuity. Finally, we conduct auxiliary RD regressions to test for differences between centers above and below the 4.5 threshold at baseline in terms of auspice (Table 2.3). Here, we find no evidence of differences in these centers for any of these outcomes.

Results

Effects of star ratings on future program quality

All of our results are correctly interpreted as intent to treat (ITT) estimates.

Star ratings. Figure 2.3 illustrates the relationship between initial ERS ratings and star ratings in five subsequent years, using binned scatter plots analogous to the first stage plots presented above. Panel (a) shows that the difference between treated and non-treated centers in likelihood of being rated 3+ stars narrowed somewhat in the first few years following the initial rating, and that

[7] The Cattaneo et al. procedure ("rddensity" in Stata) has some attractive features as a test for manipulation of the forcing variable. It does not require pre-binning of the data, and automatically adapts to the boundaries of the support of the density tests, which may be unknown. This procedure requires only the choice of bandwidth associated with the local polynomial fit. Further, the authors provide a complementary Stata command (rdbwdensity) to automatically choose this bandwidth using the median of three data-driven bandwidth selection processes.

this gap had closed completely by T+4. Note that this is likely to partially reflect a ceiling effect, in that nearly all centers in our sample were rated at least 3 stars in T+5. By contrast, when considering 4+ stars, there is no evidence of a ceiling effect and the patterns are quite similar. In particular, 3 years after the initial ERS rating, the gap in likelihood of being rated 4+ stars had closed almost completely.

Next, we formally test the magnitude and statistical significance of these relationships. Specifically, the first two rows of Table 2.4 directly correspond to Figures 2.3a & 2.3b. As the graphical evidence suggested, the substantial gaps in star ratings caused by exogenous variation in ERS scores were fully eliminated three years after ratings were assigned (T+3). We next turn to predicting the underlying components of the QRIS, to understand the type of quality improvements that drove the closing of this gap.

ERS ratings. The top panel of Table 2.4 also shows RD estimates that consider future ERS ratings as an outcome.[8] Since ERS ratings are renewed every 3 years, we would first expect to see evidence of a difference by treatment status in year T+3 (i.e. three years after treatment). Indeed, in T+3, centers below the RD threshold scored about .13 points higher than those above it, though this result is not significant. In T+4 and T+5 however, these centers scored .22 and .19 points higher, respectively, and both were statistically significant. This equates to between 40% and 50% of a standard deviation for this sample. Figure 2.4a illustrates this relationship graphically in T and T+5.

[8] Note that both ERS outcomes we consider here have mechanical results in the baseline year. Specifically, when considering average ERS ratings, the RD is unidentified because the assignment variable and outcome are the same. When considering whether a center earned above a 4.5 rating, the treatment indicator and the outcome are perfectly negatively collinear.

We also consider whether centers earned average ERS ratings above the 4.5 threshold. In year T this outcome is mechanically 0 for all treated centers and 1 for all untreated centers, reflecting the basic structure of the RD design. Yet by T+3 this gap was almost entirely closed (and no longer statistically significant), and in T+4 and T+5 treated centers were moderately *more likely* to be above this threshold (though not significantly so). Figure 2.4b illustrates this graphically.

Other measures of quality. We also examined other indicators of quality that are collected as part of SRL (not shown). In particular, we considered staff education and experience, space requirements, and staff-child ratios. We find no evidence of a difference across treated and untreated centers for any of these outcomes.

Effects of star ratings on program enrollment & closures

Enrollment. The bottom panel of Table 2.4 shows differences across treated and untreated centers in terms of enrollment and proportion of capacity filled. At baseline, there were no statistically significant differences in enrollment, though treated centers had about 5 fewer students on average than untreated centers. In T+3 and T+4, the difference across centers was marginally significant (at the .10 level). In T+5 treated centers enrolled about 11 fewer students on average than untreated centers (about 25% of a standard deviation), significant at the .05 level. Figure 2.5a depicts this relationship graphically in T and T+5. Similarly, at baseline centers on either side of the threshold did not differ with respect to the percentage of their total capacity that was filled. However, by T+5, centers below the threshold

were 8 percentage points lower along this metric, and this difference was significant. This is depicted graphically in Figure 2.5b.

Closures. In Table 2.4, we also consider two distinct forms of attrition from the sample. First, we consider whether a center was closed (i.e. no longer in operation). We find no significant difference in closure rates for treated and untreated centers. Next, we consider whether a center was still open but no longer had a valid ERS rating (i.e. chose not to be re-rated). On this outcome, we find evidence that centers below the 4.5 threshold were more likely to choose not to be re-rated than centers above it. For instance in T+3, centers just below the cut point were 10 percentage points more likely to be open but with no ERS rating, marginally significant at the .10 level. This difference across treated and untreated centers was similar in T+4 and T+5, and significant in both years. These relationships are depicted graphically at baseline and in T+5 in Figures 2.5c and 2.5d.

Robustness of the estimates

In Appendix 2D, we present results for our primary set of outcomes in T+5 using multiple alternate specifications. Here, we compare our preferred estimates to estimates that condition on a linear spline with increasingly narrow bandwidths down to the Imbens & Kalyanaraman (2009) bandwidth of 1. We also show results that use a bandwidth of 1 with a triangular kernel. Across all outcomes, the "local linear" estimates, and in particular the estimates that use the IK bandwidth, mirror our full sample estimates. This lends us confidence that our estimates are not sensitive to our choice of functional form.

Heterogeneity of results

In Table 2.5 we examine how differences between treated and untreated centers vary by enrollment and by auspice.

By enrollment. The first two columns of Table 2.5 show results for centers split by median baseline enrollment. We see considerable differences across centers with lower and higher enrollment. Among low enrollment centers, we see a large and significant difference in ERS ratings between treated and untreated centers. By comparison, we see a non-significant *negative* coefficient among high enrollment centers. In results not shown, we find no differences for either sample in terms of other measures of quality measured by SRL.

When considering enrollment, the differences we see across treated and non treated centers are concentrated among *high* enrollment centers. For instance, among high enrollment centers, those below the RD threshold had about 13 fewer children enrolled in T+5. They were also about 9 percentage points lower in terms of proportion of capacity filled. Finally, we see no differences across treated and untreated centers for either group in terms of closure rates, but lower enrollment centers below the cutoff were significantly more likely to opt not to be re-rated. For high enrollment centers, this treatment effect was not significant.

By auspice. The final two columns of Table 2.5 show how results differ when considering Head Start centers and those operating in local public schools (LPS) to other types of centers. Here, we include estimates for Head Start and LPS centers for completeness, but we caution against interpretation of these estimates. In particular, the vast majority of Head Start and LPS centers operate as part of North

Carolina's NC Pre-K program, which requires centers to maintain at least a 4 star rating at all times (NC Division of Child Development and Early Education, n.d.). This means there are very few of these types of centers that are below the 4.5 threshold, so estimating our RD models for this group is problematic. Indeed, we do not see a significant first stage for these types of when considering either 3+ or 4+ stars, indicating that these types of centers were generally not affected by the focal RD threshold of this study.

By comparison, the estimates for other center based care settings are quite similar to the results for the full sample. In particular, treated centers were no longer significantly different from untreated centers in terms of star ratings by T+5, and also had significantly higher average ERS scores than untreated centers. These centers had significantly lower enrollment relative to capacity, and were also more likely to choose not to be re-rated.

Discussion & Policy Implications

This paper examines the causal effects of receiving a lower versus higher star rating in North Carolina's Star Rated License program on subsequent program quality as well as enrollment and likelihood of closure. In exploring this specific program, we aim to shed light on whether QRIS have the potential to change quality in the early child care sector. Specifically, we examine whether two primary mechanisms of these programs operate as theorized in a fully-implemented, scaled-up statewide context.

We find evidence of multiple effects of receiving a lower star rating that are consistent with the QRIS theory of action. First, we find that centers responded to

receiving lower quality ratings by improving their scores on an observational measure of quality. They did *not*, however, improve along other measured dimensions of quality. It is important to note that this result is at least partly driven by our research design. Specifically, we leverage a treatment contrast in which treated centers stood to improve their overall star ratings by improving their ERS scores by only a small amount. By contrast, these same providers could not necessarily improve their star ratings by improving a similar amount along other dimensions. This suggests that providers responded narrowly to the incentives that they faced.

Although we treat improvements in quality as driven by providers themselves, it is possible that these improvements could also be influenced by resources provided by the state. For instance, if ERS scores can be improved simply by purchasing classroom materials or similar methods, then providers could be mechanically improving just by nature of receiving state funding. However, in North Carolina, this does not appear to be the case. In particular, North Carolina does not differentially allocate funding to providers with different star ratings, aside from differences in subsidies. Since higher rated centers receive higher subsidies, that would suggest that if anything, we should expect lower rated centers to get *worse* relative to higher rated centers. This suggests that differences in quality are in fact driven by providers.

We also find evidence of potential unintended provider responses to the QRIS. Specifically, we find that some providers who received lower ratings opted not to be re-rated on an ERS scale, which may indicate they were discouraged or

disillusioned by the rating process. This is in keeping with other literature that finds diverse effects of the same incentive structures among teachers in Washington D.C. (Dee & Wyckoff, 2015) and students in Minnesota (Dee & Jacobs, 2006). In particular, when targets are seen as easy to achieve they tend to encourage extra effort. However, when they are seen as difficult to achieve they may discourage effort.

Next, we find that centers with lower ratings lost enrollment relative to higher rated centers five years after ratings occurred, which is consistent with parents making enrollment decisions based on this information. This is also in line with Hastings & Weinstein (2008), who found that parents responded to information about quality by selectively enrolling their children in higher quality care. Notably, although we might expect to see differences in enrollment immediately after ratings were issued, we first see evidence of these differences three years after treatment. Given this result, an important caveat to mention is that differences in enrollment could be driven by either demand side responses by parents or supply side responses by providers themselves. For instance, providers that receive low ratings might choose to enroll fewer students as a way to focus more attention on each student. Since our identification strategy captures the entire effect of receiving a higher vs. lower star rating, we cannot distinguish between these multiple potential mechanisms.

We also find that responses differed across centers depending on enrollment. Specifically, both improvements in ERS ratings and the increased likelihood of not renewing an ERS rating were driven by centers with enrollment below the median,

suggesting that these lower enrollment centers were more responsive to incentives. While it is not immediately apparent that lower enrollment centers would be more responsive, this may be due to the same process of diverse responses to incentives described above. For example, it may be less costly for centers with fewer classes to improve along an observational measure of classroom quality, as these centers can focus their improvement efforts on fewer classrooms. By contrast, quality improvements in large centers may be spread across many classrooms, and thus may be less likely to be picked up by ERS ratings. However, this does not explain why higher enrollment centers who received low QRIS ratings did not opt out of the ERS rating process, as they would still face the same disincentive to be re-rated.

In contrast to the results described above, changes in enrollment were more pronounced among *high* enrollment centers. Although the differences in treatment effects between low and high enrollment centers in terms of total enrollment are striking, these groups are similar in terms of the proportion of students that were lost relative to higher rated centers. In general, differences in the effect of lower ratings on enrollment across groups may largely reflect differences in enrollment across these groups.

Limits of the current research.

Importantly, this paper does not evaluate the effect of an entire QRIS program, but rather, we evaluate the responses generated by being assigned a lower versus higher quality rating within a QRIS system. Since many such contrasts are embedded within North Carolina's QRIS, the actual effect of the entire system is likely to be

substantially higher. Future work may be able to leverage differences across states or across regimes to estimate the impact of an entire QRIS.

As discussed above, we necessarily omit a large portion of providers from our sample because they chose not to receive ERS ratings. These omitted providers are likely to be of lower average quality than those in our sample. For this reason, we might expect that provider responses to incentives may differ across providers we include and those we omit. If, for example, higher quality providers are more able to respond to incentives, these results may not generalize to the full population of care providers.

Next, although we find evidence that providers were responsive to incentives around quality, we are limited in our ability to make conclusions about how these improvements occurred and whether providers actually improved in ways that were meaningful for student learning. For instance, although ERS ratings are among the most widely used measures of quality in early childhood, recent work has raised concerns that they are not strongly related to student outcomes (e.g. Perlman, Zellman, & Le 2004; Gordon et al. 2013). Further, although we see improvement in ERS scores overall, these ratings encompass a diverse set of classroom measures, and we do not observe the specific dimensions on which these programs improved. As an example, a higher ERS rating could equate to added toys/materials, better personal care routines, more enriching interactions between children and staff, or a number of other possibilities. Some areas are likely to be easier to improve than others, and those easily-improved areas could also be less salient for student learning. This raises the concern that provider responses in North Carolina could

have been largely concentrated along easily-improved, but less important, dimensions of quality. In that case, providers may be less able to respond along other, more salient measures. Despite these concerns, these ratings are used in the majority of state QRIS, and so are highly policy relevant. As such, information about the ability of centers to respond along these dimensions is valuable.

Finally, although this paper finds evidence consistent with parents responding to information about quality in North Carolina, there are reasons to suspect this may not generalize to other states. North Carolina's program has been in operation for over 15 years, and has directed considerable time and resources toward building public awareness. In most states, QRIS systems are quite new, and are likely not to have the same degree of public awareness. Despite these caveats, this paper provides promising evidence that QRIS systems can have meaningful effects on the early child care sector in a fully-implemented, scaled-up context.

References

Araujo, M., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2014). A helping hand? Teacher quality and learning outcomes in kindergarten. *Inter-American Development Bank*.

Barnett, W. S., Friedman-Krauss, A., Gomez, R., Horowitz, M., Weisenfeld, G. G., & Squires, J. (2016). *The state of preschool 2015: State preschool yearbook*. National Institute for Early Education Research. Retrieved from http://nieer.org/sites/nieer/files/2015%20Yearbook.pdf

Bassok, D., Fitzpatrick, M., Greenberg, E., & Loeb, S. (2016). Within- and between-sector quality differences in early childhood education and care. *Child Development*, n/a-n/a. http://doi.org/10.1111/cdev.12551

Bassok, D., & Galdo, E. (2016). Inequality in preschool quality? Community-level disparities in access to high-quality learning environments. *Early Education and Development*, *27*(1), 128–144. http://doi.org/10.1080/10409289.2015.1057463

Bassok, D., & Loeb, S. (2015). Early childhood and the achievement gap. In H. F. Ladd & M. Goertz (Eds.), *Handbook of Research in Education Finance and Policy* (2nd ed.).

Boller, K., Paulsell, D., Del Grosso, P., Blair, R., Lundquist, E., Kassow, D. Z., … Raikes, A. (2015). Impacts of a child care quality rating and improvement system on child care quality. *Early Childhood Research Quarterly*, *30*, 306–315.

Campbell, F. A., Pungello, E. P., Kainz, K., Burchinal, M., Pan, Y., Wasik, B. H., … Ramey, C. T. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology*, *48*(4), 1033–1043. http://doi.org/10.1037/a0026644

Cattaneo, M. D., Jansson, M., & Ma, X. (2015). Simple local regression distribution estimators with an application to manipulation testing. *Working Paper*. Retrieved from http://www-personal.umich.edu/~cattaneo/papers/Cattaneo-Jansson-Ma_2016_LocPolDensity.pdf

Cryer, D., & Burchinal, M. (1997). Parents as child care consumers. *Early Childhood Research Quarterly*, *12*(1), 35–58.

Cryer, D., Tietze, W., & Wessels, H. (2002). Parents' perceptions of their children's child care: A cross-national comparison. *Early Childhood Research Quarterly*, *17*(2), 259–277.

Dafny, L., & Dranove, D. (2008). Do report cards tell consumers anything they don't already know? The case of Medicare HMOs. *The RAND Journal of Economics*, *39*(3), 790–821. http://doi.org/10.1111/j.1756-2171.2008.00039.x

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, *30*(3), 418–446. http://doi.org/10.1002/pam.20586

Dee, T. S., & Jacob, B. A. (2006). *Do high school exit exams influence educational attainment or labor market performance?* National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w12199

Dranove, D., Kessler, D., McClellan, M., & Satterthwaite, M. (2002). *Is more information better? The effects of "report cards" on health care providers* (Working Paper No. 8697). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w8697

Early, D., Barbarin, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R., … others. (2005). Pre-kindergarten in eleven states: NCEDL's multi-state study of pre-kindergarten & study of state-wide early education programs (SWEEP) preliminary descriptive report. *NCEDL Working Paper*. Retrieved from http://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/NCEDL_PreK-in-Eleven-States_Working-Paper_2005.pdf

Figlio, D., & Loeb, S. (2011). School Accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education. Vol. 3: [...]* (1. ed). Amsterdam: North-Holland.

Friesen, J., Javdani, M., Smith, J., & Woodcock, S. (2012). How do school "report cards" affect school choice decisions? *Canadian Journal of Economics/Revue Canadienne D'économique*, *45*(2), 784–807.

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for assessments of child care quality and its relation to child development. *Developmental Psychology*, *49*(1), 146–160. http://doi.org/10.1037/a0027899

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, *76*(5), 949–967.

Hanushek, E. A., Kain, J. F., Rivkin, S. G., & Branch, G. F. (2007). Charter school quality and parental decision making with school choice. *Journal of Public Economics*, *91*(5–6), 823–848. http://doi.org/10.1016/j.jpubeco.2006.09.014

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, *24*(2), 297–327.

Harms, T., Clifford, R., & Cryer, D. (1998). Early Childhood Environment Scale - Revised Edition.

Harms, T., Cryer, D., & Clifford, R. (2003). Infant/Toddler Environment Rating Scale - Revised Edition.

Harms, T., Cryer, D., & Clifford, R. (2007). Family Child Care Environment Rating Scale - Revised Edition.

Harms, T., Jacobs, E., & White, D. (1996). School-Age Care Environment Rating Scale.

Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics*, *123*(4), 1373–1414. http://doi.org/10.1162/qjec.2008.123.4.1373

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902.

Helburn, S. W. (1995). Cost, quality and child outcomes in child care centers. Technical report.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-Kindergarten programs. *Early Childhood Research Quarterly*, *23*(1), 27–50. http://doi.org/10.1016/j.ecresq.2007.05.002

Imbens, G., & Kalyanaraman, K. (2011). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, rdr043.

Jin, G. Z., & Leslie, P. (2002). *The effect of information on product quality: Evidence from restaurant hygiene grade cards* (SSRN Scholarly Paper No. ID 322883). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=322883

Jin, G. Z., & Sorensen, A. T. (2006). Information and consumer choice: The value of publicized health plan ratings. *Journal of Health Economics*, *25*(2), 248–275. http://doi.org/10.1016/j.jhealeco.2005.06.002

Koning, P., & van der Wiel, K. (2013). Ranking the schools: How school-quality information affects school choice in the Netherlands. *Journal of the European Economic Association*, *11*(2), 466–493. http://doi.org/10.1111/jeea.12005

Lee, D. S., & Lemieux, T. (2009). *Regression discontinuity designs in economics* (Working Paper No. 14723). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14723

Lipsey, M., Farran, D., & Hofer, K. (2015). *Evaluation of the Tennessee voluntary prekindergarten program: Kindergarten and first grade follow-up results from the randomized control design*. Nashville, TN: Peabody Research Institute. Retrieved from https://my.vanderbilt.edu/tnprekevaluation/files/2013/10/August2013_PRI_Kand1stFollowup_TN-VPK_RCT_ProjectResults_FullReport1.pdf

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., … Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732–749.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698–714. http://doi.org/10.1016/j.jeconom.2007.05.005

Mocan, N. (2007). Can consumers detect lemons? An empirical analysis of information asymmetry in the market for child care. *Journal of Population Economics*, *20*(4), 743–780.

National Center on Child Care Quality Improvement. (2015). *QRIS Resource Guide*. QRIS National Learning Network. Retrieved from https://qrisguide.acf.hhs.gov/files/QRIS_Resource_Guide_2015.pdf

National Research Council. (2011). *Incentives and test-based accountability in education*. Washington, D.C.: National Academies Press. Retrieved from http://www.nap.edu/catalog/12521

North Carolina Division of Child Development and Early Education. (n.d.). Retrieved from ncchildcare.nc.gov

North Carolina Rated License Assessment Project. (n.d.). Retrieved from www.ncrlap.org

Perlman, M., Zellman, G. L., & Le, V.-N. (2004). Examining the psychometric properties of the Early Childhood Environment Rating Scale-Revised (ECERS-R). *Early Childhood Research Quarterly*, *19*(3), 398–412. http://doi.org/10.1016/j.ecresq.2004.07.006

Peyton, V., Jacobs, A., O'Brien, M., & Roy, C. (2001). Reasons for choosing child care: Associations with family factors, quality, and satisfaction. *Early Childhood Research Quarterly*, *16*(2), 191–208.

Pope, D. G. (2009). Reacting to rankings: Evidence from "America's Best Hospitals." *Journal of Health Economics*, *28*(6), 1154–1165. http://doi.org/10.1016/j.jhealeco.2009.08.006

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., … Downer, J. (2012). Third grade follow-up to the Head Start Impact Study: Final report. OPRE report 2012-45. *Administration for Children & Families*. Retrieved from http://eric.ed.gov/?id=ED539264

QRIS National Learning Network. (2015). QRIS state contacts & map. Retrieved May 20, 2015, from http://www.qrisnetwork.org/qris-state-contacts-map

Raikes, H., Torquati, J., Wang, C., & Shjegstad, B. (2012). Parent experiences with state child care subsidy systems and their perceptions of choice and quality in care selected. *Early Education & Development*, *23*(4), 558–582.

Rockoff, J. E., & Turner, L. J. (2010). *Short Run Impacts of Accountability on School Quality* (SSRN Scholarly Paper No. ID 1949323). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1949323

Sabol, T. J., Hong, S. S., Pianta, R. C., & Burchinal, M. (2013). Can rating pre-k programs predict children's learning? *Science*, *341*(6148), 845–846.

Sabol, T. J., & Pianta, R. C. (2014). Do standard measures of preschool quality used in statewide policy predict school readiness? *Education Finance and Policy*, *9*(2), 116–164.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). *Standards for regression discontinuity designs*. What Works Clearinghouse. Retrieved from http://eric.ed.gov/?id=ED510742

Schweinhart, L., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry preschool study through age 40*. Ypsilanti: High Scope Press.

Snow, C., & Van Hemel, S. (2008). *Early childhood assessment: Why, what, and how?* National Research Council of the National Academies.

The Build Initiative & Child Trends. (2015). A catalog and comparison of Quality Rating and Improvement Systems (QRIS) [Data system]. Retrieved from www.qriscompendium.org

Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). Issues for the next decade of quality rating and improvement systems. *Washington, DC: Child Trends*. Retrieved from http://www.acf.hhs.gov/sites/default/files/opre/next_decade.pdf

Van Horn, M. L., Ramey, S. L., Mulvihill, B. A., & Newell, W. Y. (2001). Reasons for child care choice and appraisal among low-income mothers. In *Child and Youth Care Forum* (Vol. 30, pp. 231–249). Springer. Retrieved from http://link.springer.com/article/10.1023/A:1016755630684

Wong, M., Cook, T. D., & Steiner, P. M. (2015). Adding design elements to improve time series designs: No Child Left Behind as an example of causal pattern-matching. *Journal of Research on Educational Effectiveness*, *8*(2), 245–279. http://doi.org/10.1080/19345747.2013.878011

Wrigley, J., & Dreby, J. (2005). Fatalities and the organization of child care in the United States, 1985-2003. *American Sociological Review*, *70*(5), 729–757.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Espinosa, L. M., Gormley, W. T., … Zaslow, M. J. (2013). Investing in our future: The evidence base on preschool education.

Zellman, G. L., & Perlman, M. (2008). *Child-care quality rating and improvement systems in five pioneer states*. Rand Corporation. Retrieved from http://www.rand.org/pubs/monographs/MG795/

Figure 2.1 First stage relationships between average ERS ratings and star ratings at baseline

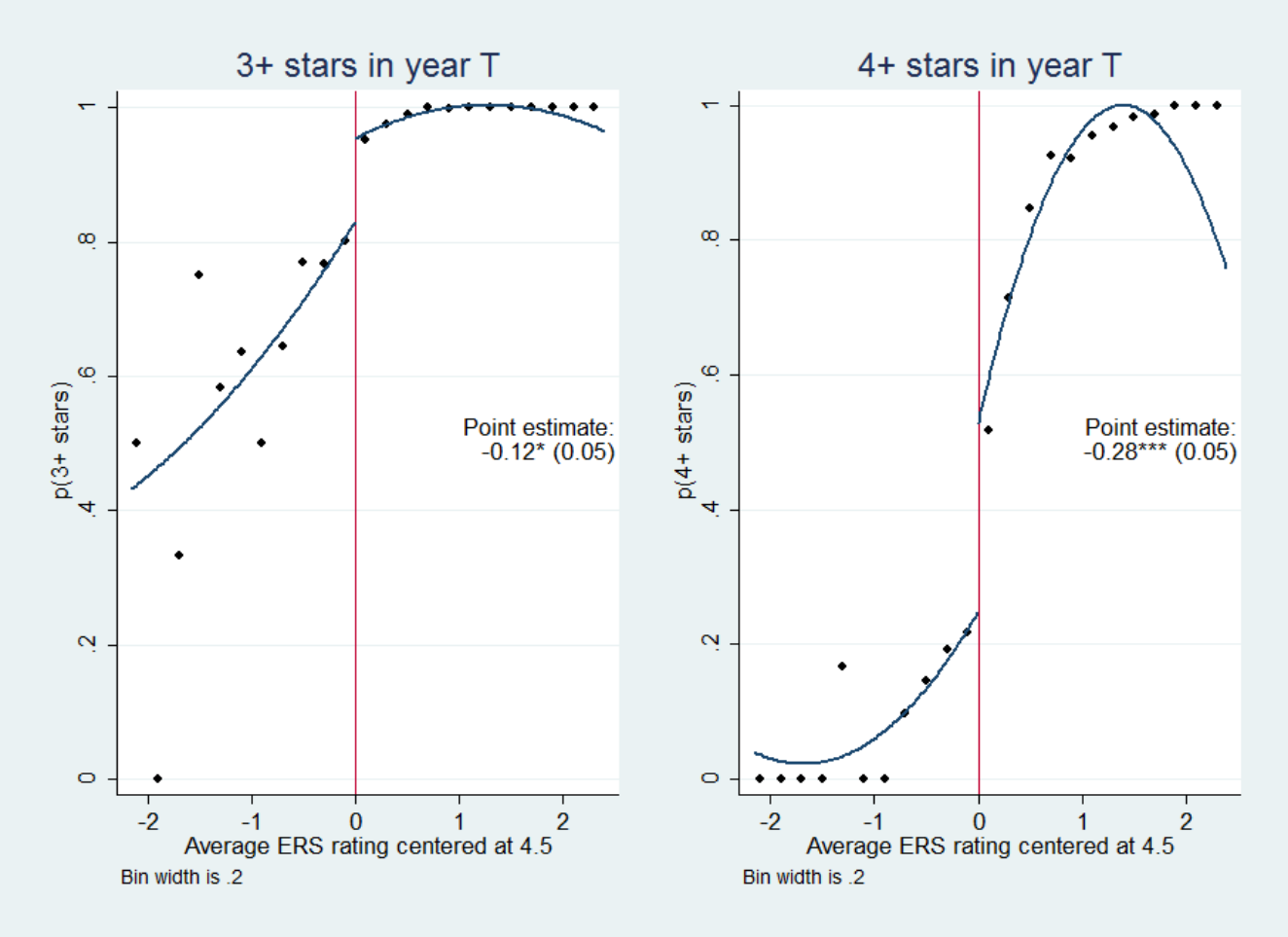
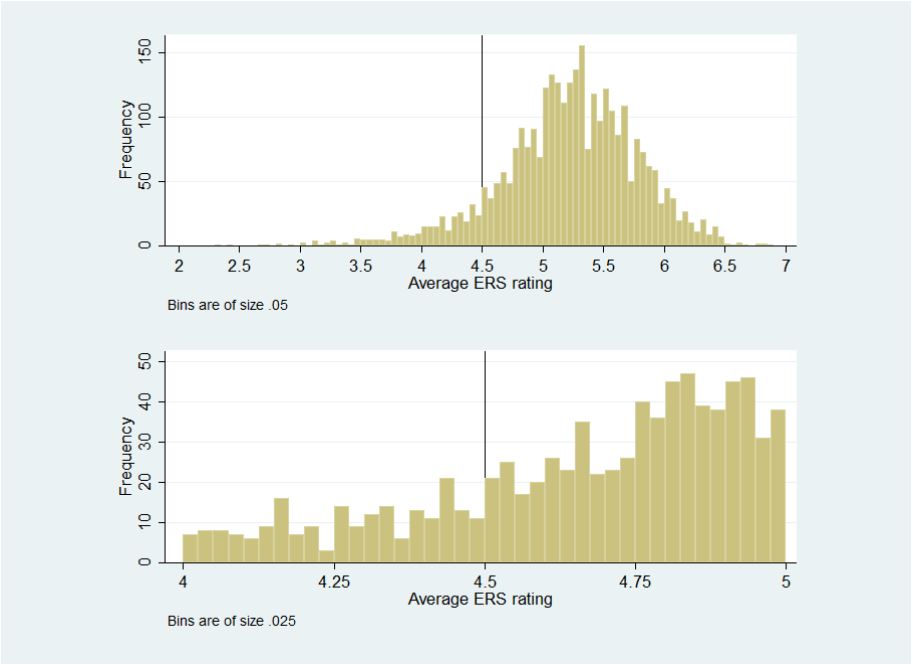


Figure 2.2 Density of forcing variable around RD threshold

(1) Binned histograms of average ERS rating using bin widths of .05 and .025



(2) McCrary density test for discontinuity at the 4.5 threshold

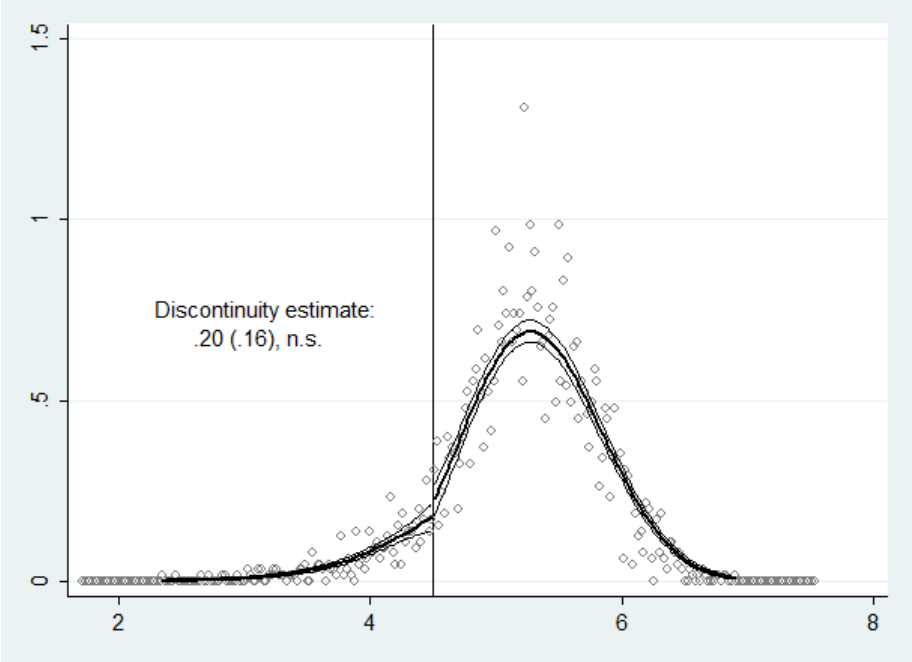
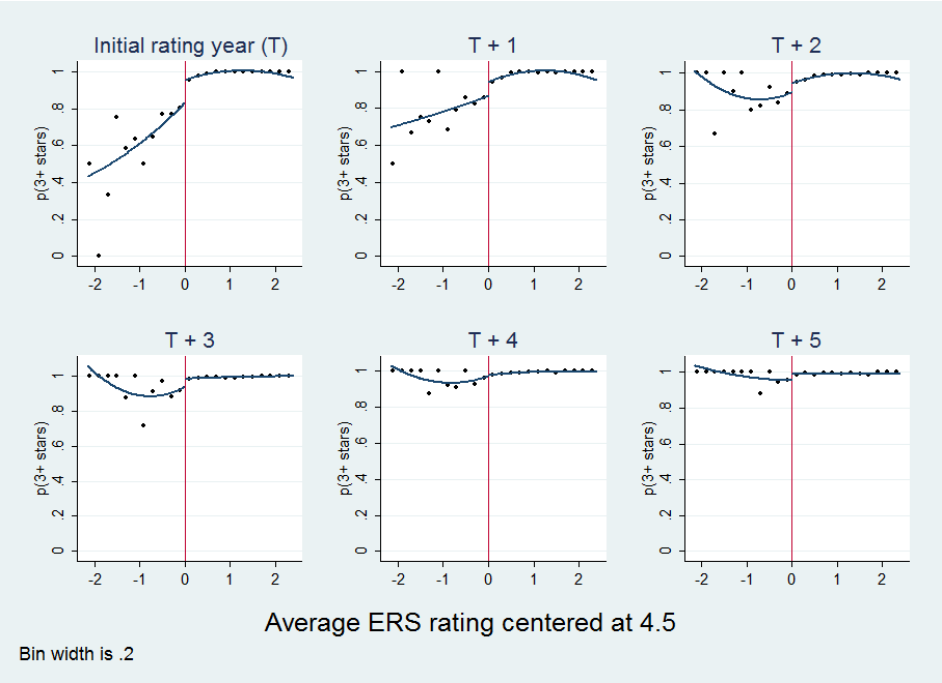


Figure 2.3 Reduced form relationships between average ERS ratings at baseline (T) and star ratings in subsequent years

(a) 3+ stars



(b) 4+ stars

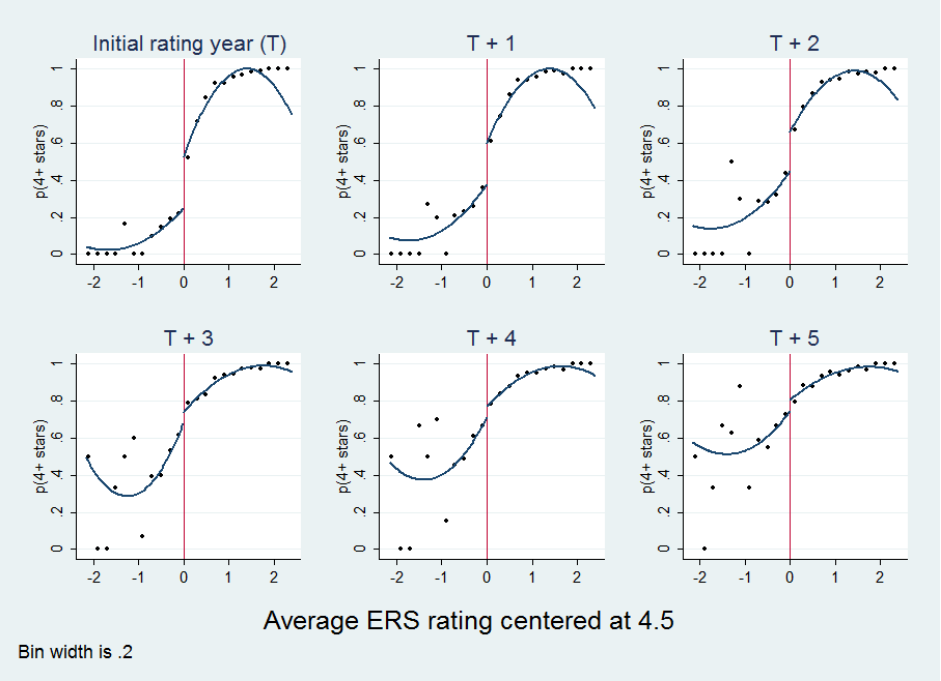
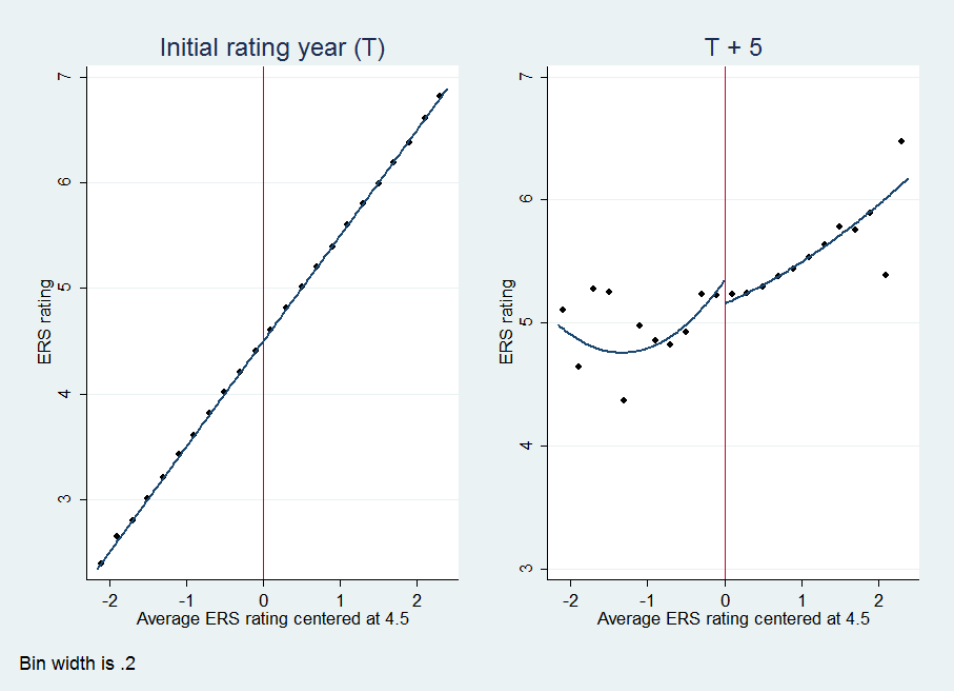


Figure 2.4 Reduced form relationships between average ERS ratings at baseline and in T+5

(a) Average ERS rating



(b) Average ERS score ≥ 4.5

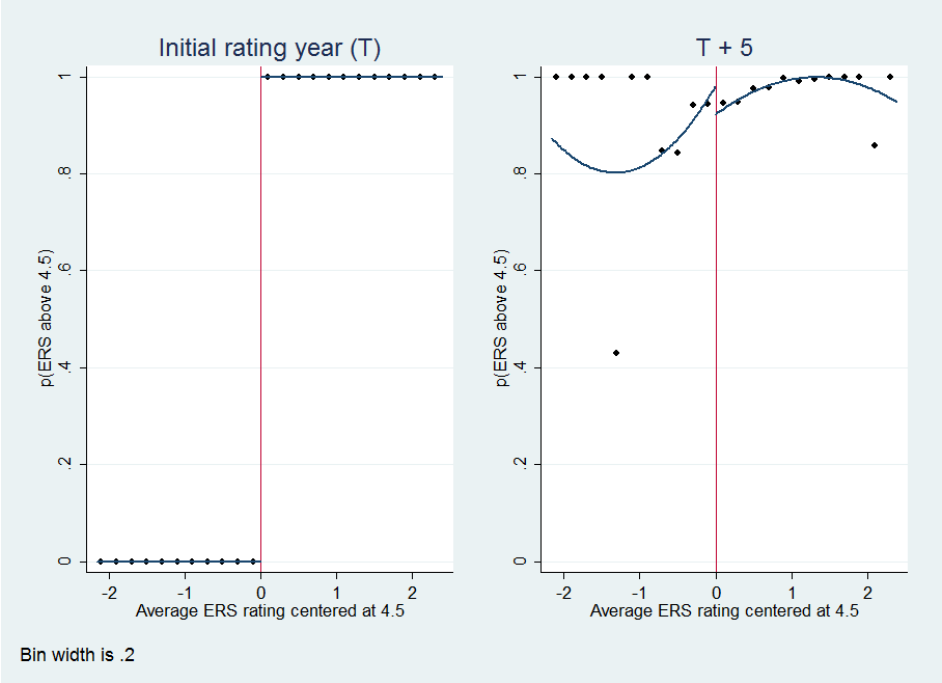
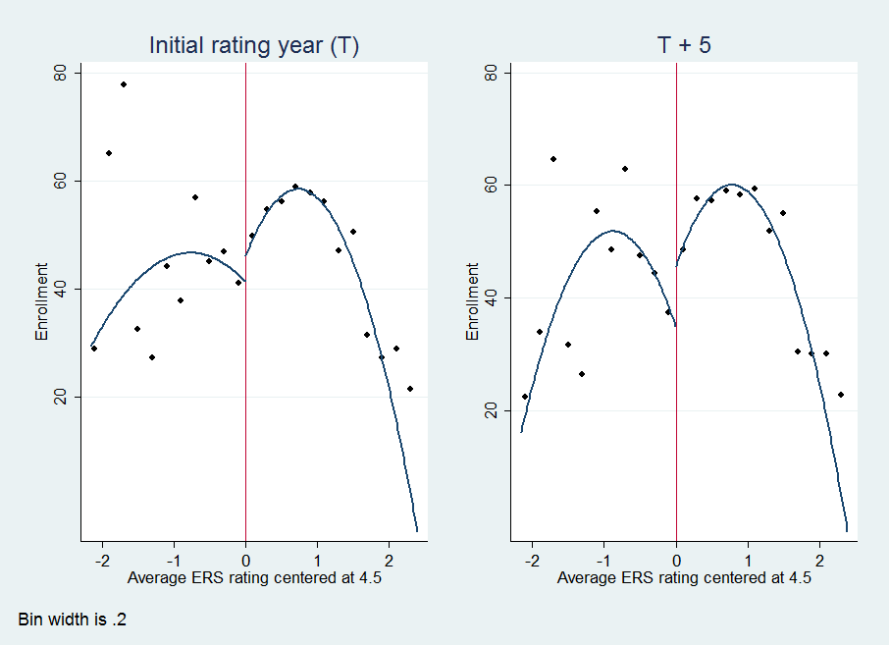
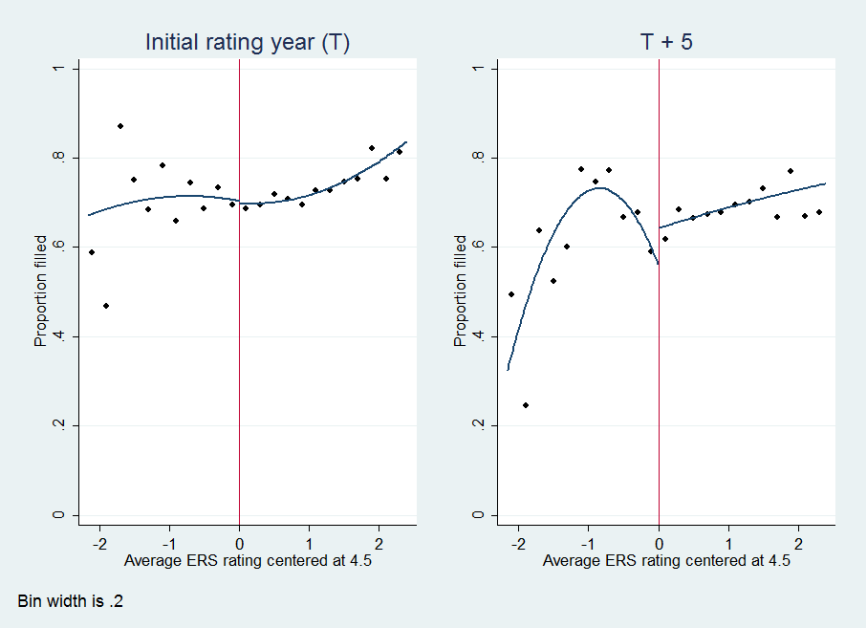


Figure 2.5 Reduced form relationship between average ERS ratings at baseline and enrollment/closures

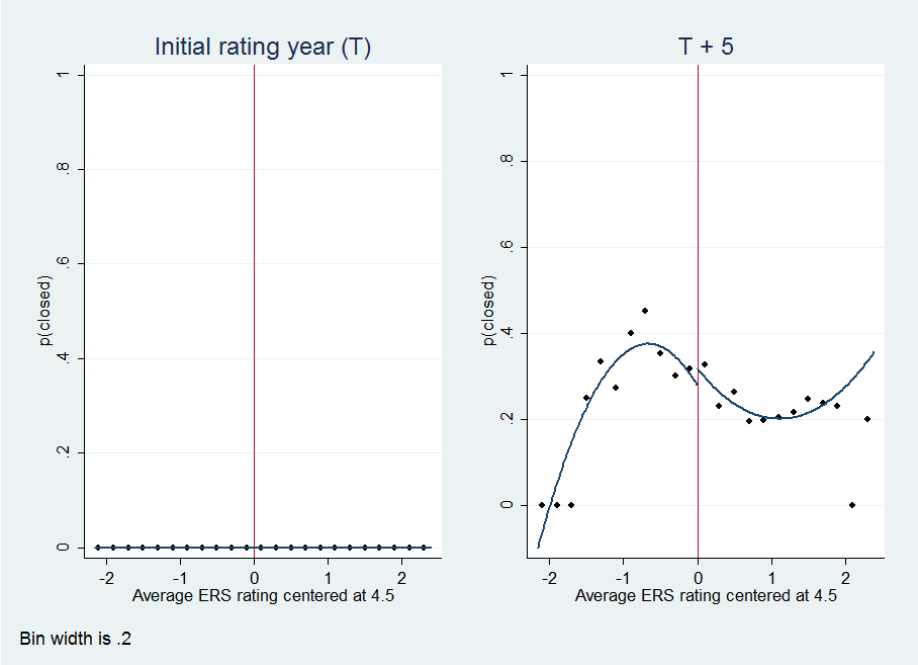
(a) Enrollment



(b) Proportion of capacity filled



(c) Center was closed



(d) Center was open but with no ERS rating

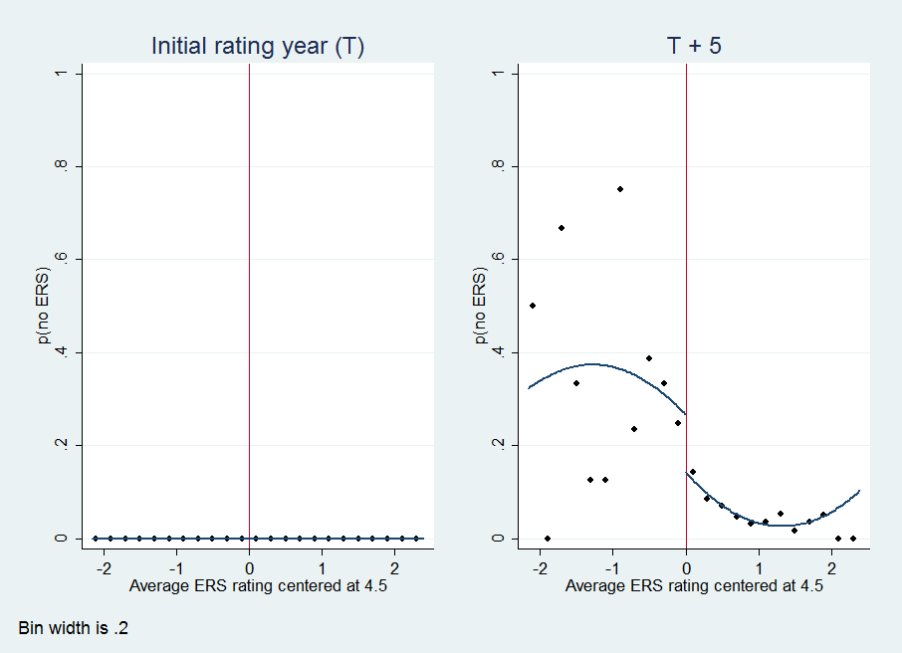


Table 2.1. Descriptive statistics for the analytic sample at baseline (T) through T+5

	T	T+1	T+2	T+3	T+4	T+5
Center was closed (i.e. not in operation)	0.00	0.05	0.11	0.16	0.20	0.24
Conditional on being open						
Independent center	0.45	0.44	0.45	0.44	0.43	0.42
Local public school	0.27	0.27	0.26	0.27	0.27	0.28
Head Start	0.09	0.09	0.09	0.09	0.09	0.09
Religious sponsored	0.08	0.08	0.08	0.08	0.08	0.08
Other center based care	0.12	0.12	0.12	0.12	0.12	0.12
3+ star rating	0.97	0.97	0.98	0.98	0.99	0.99
4+ star rating	0.81	0.84	0.85	0.87	0.89	0.90
5 star rating	0.44	0.47	0.49	0.55	0.59	0.61
No Environment Rating Scale (ERS) rating	0.00	0.02	0.03	0.06	0.07	0.08
Average ERS rating	5.21	5.23	5.26	5.36	5.40	5.43
Average ERS rating ≥ 4.5	0.90	0.92	0.93	0.95	0.97	0.98
Capacity	80.0	81.7	83.7	84.7	85.9	86.6
Total enrollment	53.0	54.3	54.1	53.8	54.9	54.6
Proportion of capacity filled	0.71	0.71	0.70	0.69	0.69	0.68
N (conditional on being open)	3157	2990	2807	2659	2519	2411

Table 2.2 First stage estimates across specifications and bandwidth restrictions

	Quadratic		Linear			
Outcome	Full sample	Full sample	1.5	1.25	1	Triangular kernel
3+ stars	-0.12*	-0.15***	-0.15***	-0.13**	-0.12**	-0.14**
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
4+ stars	-0.28***	-0.46***	-0.42***	-0.37***	-0.32***	-0.27***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
N	3157	3157	2948	2617	2143	2120

Note. Each coefficient represents the results from a separate regression discontinuity estimate. Robust standard errors in parenthesis.
+ p < .10 * p < .05 ** p < .01 *** p < .001

Table 2.3 Auxiliary RD regressions of baseline covariate balance in terms of auspice

Independent center	-0.02
	(0.06)
Local public school	-0.01
	(0.04)
Head Start	0.04
	(0.04)
Religious sponsored	-0.03
	(0.03)
Other center based care	0.02
	(0.04)
N	3157

Note. Each coefficient represents the results from a separate regression discontinuity estimate. Each estimate conditions on a quadratic spline of the assignment variable as well as an indicator equal to 1 if a center scored below the RD threshold. Robust standard errors in parenthesis.
+ p < .10 * p < .05 ** p < .01 *** p < .001

Table 2.4 Reduced-form RD estimates for outcomes at baseline (T) through T+5

Outcome	T	T+1	T+2	T+3	T+4	T+5
Program quality						
3+ stars	-0.12*	-0.08+	-0.05	-0.04	-0.00	-0.04
	(0.05)	(0.04)	(0.04)	(0.03)	(0.02)	(0.03)
4+ stars	-0.28***	-0.22***	-0.21***	-0.06	-0.06	-0.06
	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Average ERS rating	-	0.02	0.01	0.13	0.22*	0.19*
	-	(0.04)	(0.06)	(0.10)	(0.09)	(0.08)
Average ERS ≥ 4.5	-1	-0.81***	-0.66***	-0.07	0.06	0.06
	-	(0.05)	(0.06)	(0.07)	(0.05)	(0.04)
Enrollment & closure						
Total enrollment	-4.80	-3.94	-4.68	-8.59+	-7.92+	-10.80*
	(4.45)	(4.25)	(4.53)	(4.66)	(4.64)	(4.68)
Proportion of capacity filled	0.01	0.01	0.02	-0.04	-0.02	-0.08*
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Center is closed	-	-0.00	-0.03	-0.06	-0.07	-0.04
	-	(0.03)	(0.04)	(0.05)	(0.05)	(0.06)
Open, no ERS rating	-	-0.01	0.01	0.10+	0.13*	0.12+
	-	(0.03)	(0.04)	(0.06)	(0.06)	(0.07)
N	3157	2990	2807	2659	2519	2411

Note. Each coefficient represents the results from a separate regression discontinuity estimate. Each estimate conditions on a quadratic spline of the assignment variable as well as an indicator equal to 1 if a center scored below the RD threshold. Robust standard errors in parenthesis.

+ p < .10 * p < .05 ** p < .01 *** p < .001

Table 2.5 Heterogeneity of results by enrollment and auspice

	Low enrollment centers		High enrollment centers		Head Start/Local public school		Non Head Start/LPS	
Outcome	T	T+5	T	T+5	T	T+5	T	T+5
Program quality								
3+ stars	-0.14*	0.01	-0.09	-0.09+	-0.03	0.00***	-0.15**	-0.05
	(0.06)	(0.01)	(0.07)	(0.05)	(0.12)	(0.00)	(0.06)	(0.04)
4+ stars	-0.22**	0.01	-0.35***	-0.12	-0.12	0.00	-0.27***	-0.08
	(0.07)	(0.09)	(0.08)	(0.10)	(0.16)	(0.12)	(0.05)	(0.08)
Average ERS rating	-	0.52***	-	-0.10	-	-0.13	-	0.15+
	-	(0.13)	-	(0.10)	-	(0.28)	-	(0.08)
Average ERS ≥ 4.5	-1.00	0.14+	-1.00	0.01	-1.00	-0.31+	-1.00	0.12***
	-	(0.08)	-	(0.04)	-	(0.18)	-	(0.03)
Enrollment & closure								
Total enrollment	2.70+	1.26	0.63	-13.36*	-30.09***	-24.02**	3.40	-3.91
	(1.64)	(2.62)	(6.93)	(6.81)	(8.22)	(7.82)	(5.33)	(5.85)
Proportion of capacity filled	0.00	-0.08	0.03	-0.09+	-0.16	-0.13	0.03	-0.09*
	(0.05)	(0.06)	(0.04)	(0.05)	(0.11)	(0.12)	(0.04)	(0.04)
Center is closed	-	-0.04	-	-0.10	-	0.08	-	-0.06
	-	(0.08)	-	(0.07)	-	(0.15)	-	(0.07)
Open, no ERS rating	-	0.19*	-	0.04	-	-0.04	-	0.21*
	-	(0.09)	-	(0.09)	-	(0.09)	-	(0.08)
N	1575	1091	1582	1320	1125	897	2030	1512

Note. Each coefficient represents the results from a separate regression discontinuity estimate. Each estimate conditions on a quadratic spline of the assignment variable as well as an indicator equal to 1 if a center scored below the RD threshold. Robust standard errors in parenthesis. Low enrollment and high enrollment are defined as centers below and above the median baseline enrollment of 40.

+ p < .10 * p < .05 ** p < .01 *** p < .001

Appendix 2A. Calculation of program standards scores in North Carolina

In North Carolina, the program standards component of the QRIS accounts for nearly half of the total points that centers can receive (i.e. 7 out of a total 15). Criteria for the program standards component build on one another so that to receive a higher score a center must meet all requirements for each of the lower scores. Specifically, points are earned as follows. Many of these requirements refer to "enhanced standards," which are detailed in full immediately afterward.

Program standards score	Requirement
1	Meets minimum licensing requirements
2	Meets all enhanced standards except either staff-child ratios OR space requirements
3	Lowest classroom ERS score ≥ 4.0
4	Meets all enhanced standards except space requirements AND average ERS score ≥ 4.5 with no single score below 4.0
5	Average ERS score ≥ 4.75 with no single score below 4.0
6	Meets all enhanced standards AND average ERS score ≥5.0 with no single score below 4.0
7	Meets enhanced ratios minus 1 AND lowest classroom ERS score ≥ 5.0

Enhanced program standards (North Carolina Division of Child Development 2009):

Space requirements

- There must be at least 30 sq ft of inside space and 100 sq ft outside space per child per the licensed capacity, OR
- There must be at least 35 sq ft of inside space and 50 sq ft outside space per child per the licensed capacity
- There must be an area which can be arranged for administrative and private conference activities

Staff child ratios

- Staff-child ratios must be posted at all times in a prominent classroom area
- To meet enhance staff child ratio requirements, centers must meet the following criteria:

Age of children served	Staff child ratio	Maximum group size
0-12 months	1/5	10
1-2 years	1/6	12
2-3 years	1/9	18
3-4 years	1/10	20

Administrative policies:

- Selection and training of staff
- Communication with and opportunities for participation by parents
- Operational and fiscal management
- Objective evaluation of the program, management, and staff

Personnel policies

- Each center with 2 or more staff must have written personnel policies including job descriptions, minimum qualifications, health & medical requirements etc.
- Personnel policies must be discussed with each employee at the time of employment and copies must be available to staff
- Each employee's personnel file must contain an annual evaluation and development plan
- Personnel files must contain a signed statement verifying that the employee has received and reviewed personnel policies

Operational policies

- Must have written policies that describe the operation of the center and services which are available to children/parents, including days/hours of operation, age range of children served, parent fees, etc.
- Operational policies must be discussed with parents when they inquire about enrolling their child, and written copies must be provided
- Copies of operational policies must be distributed to all staff

Caregiving activities for preschool aged children

- Each center providing care to preschool-age children 2 or older must provide all five of the following activity areas daily
    - Art/creative play
    - Children's books
    - Block & block building
    - Manipulatives
    - Family living & dramatic play
- The following activities must also be provided at least once per week
    - Music and rhythm
    - Science and nature
    - Sand/water play

Parent participation

- Each center must have a plan to encourage parent participation and inform parents about programs/services that includes the following
    - A procedure for encouraging parents to visit the center before their child starts attending
    - Opportunities for staff to meet with parents on a regular basis
    - Activities which provide parents opportunities to participate
    - A procedure for parents who need information or have complaints about the program
- The plan must be provided to and discussed with parents when the child is enrolled

Appendix 2B. Sample five star rated license

State of North Carolina
Department of Health and Human Services
Division of Child Development and Early Education

Five Star Child Care License

[redacted]

In each area rated, this facility earned:

Staff Education: 6 out of 7 points
Program Standards: 6 out of 7 points
Quality Point: 1 out of 1 points

Education Option Met: ☒ Programmatic Option Met: ☐

Total: 13 out of 15 points

ID Number: 11000614
Type of Facility: Center
Issued to: [redacted]

Age Range: 0 - 12 years
Capacity: 1st shift: 75; 2nd shift: 0; 3rd shift: 0
Effective Date: March 1, 2012
Restrictions:
Daytime care only
Meets enhanced ratios
Meets enhanced space

In accordance with Article 7, Chapter 110 of the North Carolina General Statutes, the above named child care facility is issued a rated license. Licenses vary from an overall rating of one through five stars, based upon their cumulative points in the three categories above.

This license must be displayed in a prominent place so it may be available and shown to each child's parent or guardian when the child is enrolled. This license cannot be bought, sold or transferred. It is valid only for the location/address noted above. This license is the property of the State of North Carolina and must be returned to the Division of Child Development and Early Education in the event of termination or revocation.

Lanier M. Cansler, Secretary, Department of Health and Human Services

Deborah J. Cassidy, Director, Division of Child Development and Early Education

Appendix 2C. Comparison of sample and nonsample providers 2007-2009

	2007		2008		2009	
	Sample	Nonsample	Sample	Nonsample	Sample	Nonsample
Independent center	0.44	0.53***	0.44	0.52***	0.45	0.53***
Local public school	0.27	0.17***	0.27	0.17***	0.27	0.16***
Head Start	0.10	0.01***	0.09	0.01***	0.09	0.02***
Religious sponsored	0.08	0.21***	0.08	0.22***	0.08	0.22***
Other center based care	0.12	0.07***	0.12	0.07***	0.12	0.07***
Capacity	79.22	72.09***	80.57	73.04***	81.95	73.04***
Enrollment	54.22	43.41***	53.42	42.80***	53.36	40.42***
Proportion of capacity fille	0.73	0.64***	0.71	0.62***	0.71	0.59***
3+ star rating	0.92	0.43***	0.92	0.34***	0.97	0.37***
4+ star rating	0.73	0.10***	0.76	0.07***	0.83	0.07***
5 star rating	0.38	0.01***	0.42	0.00***	0.46	0.00***
N	2970	2050	3053	1977	2952	2000

Note. + p < .10 * p < .05 ** p < .01 *** p < .001

Appendix 2D. Sensitivity of results in year T+5 to alternate specifications and bandwidths

	Quadratic		Linear			
Outcome	Full sample	Full sample	1.5	1.25	1 (IK bandwidth)	Triangular kernel
Program quality						
3+ stars	-0.04	-0.04*	-0.04*	-0.04+	-0.04	-0.03
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
4+ stars	-0.06	-0.14**	-0.14**	-0.11*	-0.05	-0.04
	(0.07)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)
Average ERS rating	0.19*	0.14*	0.18*	0.16*	0.16*	0.14+
	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)
Average ERS ≥ 4.5	0.06	-0.00	0.03	0.01	0.02	0.04
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Enrollment & closure						
Total enrollment	-10.80*	-19.42***	-15.79***	-16.83***	-16.15***	-13.48**
	(4.68)	(3.64)	(3.80)	(4.08)	(4.46)	(4.55)
Proportion of capacity filled	-0.08*	-0.01	-0.03	-0.06+	-0.07*	-0.07+
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Center is closed	-0.04	0.09*	0.05	0.01	-0.01	-0.02
	(0.06)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)
Open, no ERS rating	0.12+	0.19***	0.20***	0.16**	0.09	0.10
	(0.07)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
N	2411	2411	2252	1996	1618	1601

Note. Each coefficient represents the results from a separate regression discontinuity estimate. Each estimate conditions on a quadratic spline of the assignment variable as well as an indicator equal to 1 if a center scored below the RD threshold. Robust standard errors in parenthesis.
+ p < .10 * p < .05 ** p < .01 *** p < .001

Chapter 3 - Examining the gap in preschool enrollment between Hispanic and non-Hispanic children

Scott Latham

Abstract

Access to high quality preschool may provide a promising way to close achievement gaps between white and Hispanic children, which are already apparent by the time children arrive in kindergarten. However research has consistently documented that Hispanic children are enrolled in center-based care at lower rates than their non-Hispanic peers. Despite the potential benefits of increasing preschool enrollment among Hispanic children, we have an underdeveloped understanding of *why* Hispanic children attend at different rates, and how patterns of enrollment differ *among* Hispanic children, questions that are essential for designing effective policy solutions.

This paper leverages new, uniquely detailed data from a nationally representative sample of households with children under age 13 to provide a portrait of how patterns of child care use differ across race/ethnicity and across Hispanic subgroups both for infants/toddlers (0-2) and for older children (3-5). I find substantial Hispanic-white gaps in preschool enrollment and show that these gaps are largest among immigrant and non-English speaking households. I demonstrate that differences in income and socioeconomic status across groups are associated with enrollment gaps are almost fully explained by differences in income and socioeconomic status across groups. Among infants and toddlers, Hispanic children attend preschool at lower rates than white children even after accounting for an extensive set of child and family characteristics.

The achievement gap between white and Hispanic children is already sizeable by the time they arrive at kindergarten. Almost a quarter of children in the United States are Hispanic, and these children score about .5 to .75 standard deviations lower on math and reading tests at school entry relative to their white counterparts (O'Hare, 2011; Reardon, Robinson-Cimpian, & Weathers, 2015). These early gaps have lasting consequences, as early academic skills are predictive of outcomes well into the future (Duncan et al., 2007; Chetty et al., 2011). Large gaps between white and Hispanic students persist as children progress through school and beyond, and are evident not only in test scores but also in rates of college completion, choice of occupation, wages, and total income (Black et al. 2006; U.S. Census Bureau, 2011, 2012, 2015).

Early childhood interventions, including preschool[9], have become increasingly popular policy levers aimed at reducing these early disparities. In particular, a large and consistent body of evidence has shown that these programs have the potential to boost school readiness among young children, and that benefits are often most pronounced among low-income and minority children (e.g. Magnuson & Waldfogel, 2005; Schweinhart et al., 2005; Heckman, 2006; Deming, 2009; Bassok, 2010; Camilli, Vargas, Ryan, & Barnett, 2010; Weiland & Yoshikawa, 2013).

Research using statewide and national datasets has found that Hispanic children in particular experience disproportionate benefits from high quality early

[9] Throughout this paper, I use the terms "preschool" and "center-based care" interchangeably to refer to any child care provided in a nonresidential group setting including public preschool and Head Start programs as well as programs operated in local public schools.

care (Gormley & Phillips, 2005; Loeb et al., 2007; Bloom & Weiland, 2015). One potential explanation is that Hispanic children receive additional benefits from extra exposure to English. For instance, Bloom & Weiland (2015) found larger effects of Head Start among Spanish speakers and dual language learners. Other work has found that Head Start raises English language proficiency among children of immigrants (Magnuson, Lahaie, & Waldfogel, 2006).

Despite this, research across two decades has found that Hispanic children are consistently enrolled in preschool at lower rates than other children (e.g. West, Germino Hausken, & Collins, 1993; Fuller et. al., 1994; Liang, Fuller, & Singer, 2000; Early & Burchinal, 2001; Bridges et al., 2004; Fuller & Kim, 2011). In 2014 about 32% of Hispanic children 3-5 were enrolled in preschool, compared with 41% of white children, a 9 point gap (Kena et al., 2016). It is encouraging to note that this gap has narrowed somewhat since 1991 (West et al., 1993). However, the enrollment gap is still sizable and may contribute to differences in achievement between white and Hispanic children, especially given that Hispanic children may benefit disproportionately from center-based care.

While researchers and practitioners have posited that increasing preschool take up among Hispanic children would yield benefits, we have an under-developed understanding of exactly *why* Hispanic children attend preschool at lower rates. Previous research has identified family characteristics related to gaps in enrollment between Hispanic and white children, including income/SES, household composition, and parent work schedules. However, black children, who have similar family characteristics to Hispanic children, have been found to attend preschool at

rates equal to or higher than white children (Fuller et al., 1996; Early & Burchinal, 2001; Radey & Brewster, 2007). This has led researchers to suggest that Hispanic families may differ from non-Hispanic families in terms of their cultural preferences and practices regarding child care (Becerra & Chi, 1992; Fuller et. al, 1994; Buriel & Hurtado-Ortiz, 2000; Harwood et al., 2002).

However, research exploring gaps in preschool enrollment has generally treated Hispanic children as a single group (Fuller et al. 1994; Early & Burchinal, 2001; Radey & Brewster, 2007), when Hispanic households are quite diverse in terms of home language, immigrant status, and country of origin (Logan & Turner, 2013; Turner et al., 2015). There is little research that examines enrollment gaps separately among Hispanic subgroups, and in fact a recent report explicitly stated the need for research detailing variation in child care use among Hispanic children by characteristics such as home language (Mendez, Crosby, & Helms, 2016).

Further, research in this area has generally focused on children of preschool age (3-5) (Fuller et al, 1994; Fuller et al., 1996; Liang et al, 2001; Hirshberg, Huang, & Fuller, 2005), with only a few studies examining enrollment in preschool for younger children (e.g. Early & Burchinal, 2001; Radey & Brewster, 2007). No studies have explicitly documented how enrollment gaps vary by age for Hispanic subgroups. These differences by age are worth exploring for a few reasons. First, given that preschool is a more normative child care option among 3-5 year olds, family characteristics may play less of a role in determining whether children of this age attend (Coley et al., 2014). By contrast, children 0-2 are much less likely to be enrolled in preschool care, and so family characteristics may play a larger role in

determining whether a child of this age is enrolled. Indeed, Coley et al. (2014) found that race/ethnicity was more strongly predictive of early child care use among infants than among preschool aged children. Next, there is increasing policy interest around expanding access to preschool for the youngest children. For example, Early Head Start, which is targeted at low-income children 3 years and younger, has expanded rapidly in recent decades (Vogel et al., 2006). As there is more and more interest in access for these younger children, it is important to document differences in access across groups.

Importantly, children 0-2 are likely to have different child care needs than preschool age children. Although evidence strongly suggests benefits to attending preschool among 3-5 year olds, the evidence is less clear for infants and toddlers. Some studies have found positive effects of preschool care for these children across a range of outcomes including cognition, language development, learning competency, and reduced aggression (Burchinal et al., 2000; Love et al, 2003), while others have found evidence of negative effects on social adjustment, maternal attachment, and stress responses (Phillips & Adams, 2001; Love et al, 2003, Watamura et al., 2003). For this reason, parents who differ in their preferences and beliefs around child rearing may choose different types of care for these young children. For instance, parents who place higher value on nurturing attachment and social development among young children may choose not to enroll children in preschool care until they are a bit older.

Another limitation of the extant literature is that most examinations of Hispanic preschool enrollment patterns are relatively old (e.g. Fuller et al, 1994;

Fuller et al., 1996; Liang et al, 2001). There have been significant changes to both the early childhood landscape and the population of Hispanic children in the United States since these questions were last examined at a national level (U.S. Census Bureau, 2011; Barnett et al, 2016; Bassok & Latham, 2016), and these changes may have influenced child care use across groups. An updated look at differences in preschool enrollment between all Hispanic children and non-Hispanic children is thus warranted. The current study leverages newly available data to document how patterns of child care use vary between Hispanic and white children, and explore how differences in care use are related to household characteristics. This data was collected in 2012, and unlike previous national surveys, it contains information about all care providers for all young children in each household. Specifically, I address the following research questions:

1. How large are gaps in non-parental care and preschool use between Hispanic children and other children?
2. To what extent do patterns vary for infants and toddlers (age 0-2) relative to preschool age children (3-5, not in K)?
3. Do patterns differ *within* Hispanic children, by home language or by immigration status?
4. To what extent are enrollment gaps – both overall and for Hispanic subgroups – related to household income/socioeconomic status, household composition, and parent work schedules?

In answering these questions, I aim to provide an updated portrait of how family characteristics are related to differences in preschool enrollment rates.

Background

Hispanic-white achievement gaps have become increasingly important in light of changing demographics in the United States (Todd & Wolpin, 2007; Reardon & Galindo, 2009; Hemphill & Vanneman, 2011; Reardon et al., 2015). Hispanics became the largest minority population in 2004, and as of 2014 comprised 17.4 percent of the total U.S. population (U.S. Census Bureau, 2014). Hispanics also make up the fastest growing segment of the population, accounting for over half of the total population growth between 2000 and 2010. In those years, the Hispanic population increased by 43%, compared just over 1% for whites (US Census Bureau, 2010). As a share of the population, the number of Hispanic children is growing even faster, and in 2014 they accounted for 24% of all children (Federal Interagency Forum on Child and Family Statistics, 2015). From both an equity and an economic perspective, the gap in achievement between white and Hispanic students has pronounced consequences.

In response to these gaps, and following a large body of evidence that documents the benefits of preschool among low income and minority children (e.g. Magnuson & Waldfogel, 2005; Camilli et al., 2010; Ansari & Winsler, 2011), public investment in early childhood education has increased dramatically. Between 2002 and 2015, state spending on preschool initiatives more than doubled from $2.4 to $6.2 billion, and over the past two decades the number of students in public preschool has nearly doubled as well. Most of these public preschool programs explicitly target underserved or disadvantaged populations, including a disproportionate number of Hispanic children (Barnett et al., 2016).

Within a similar time frame, the gap in preschool enrollment between Hispanic and white children has narrowed somewhat. In 1991, Hispanic children were 15 percentage points less likely to attend care relative to white children (West et al, 1993). This gap was down to 10 percentage points by 2005, and 9 points in 2014 (Chernoff et al., 2007; Kena et al, 2016). No research has established this relationship as causal, but it is consistent with the hypothesis that increases in targeted public preschool have reduced disparities in use of preschool between Hispanic and white children.

Despite encouraging evidence that gaps in preschool enrollment have narrowed in recent years, they are still large enough to potentially exacerbate disparities between Hispanic and white children. As such, a large body of research has attempted to understand what contributes to these gaps, with the goal of informing policy. Specifically, researchers have pointed to at least 5 sets of potential explanations, including (1) income/SES (2) household composition (3) parent work schedules (4) parental preferences/practices (5) aspects of the child care supply. In this paper I focus on the first three.

Below, I summarize research that has explored how each set of explanations relates to differences in child care selection by race/ethnicity. In practice, the five potential explanations may each contribute, and they are certainly correlated and inter-related.

Factors that contribute to parent selection of early care

One way to understand gaps in child care use across groups is using a "constraints and preferences" framework (Meyers and Jordan, 2006; Radey & Brewster, 2007;

Coley, Votruba-Drzal, Collins, & Miller, 2014). This framework asserts that decisions about child care are the result of a combination of family demands, social/cultural expectations, and constraints on information/resources. This work provides theoretical context to help understand why the following characteristics of households and their environment are related to child care decisions.

Income/SES. One commonly cited factor related to lower use of child care among Hispanic families is lower levels of household income and other measures of SES. From a constraints and preferences standpoint, extra income and resources allow parents to choose from a larger and generally higher quality set of options. Research has consistently found that families with higher income are more likely to use non parental child care in general, and center-based care specifically (Fuller et al., 1994; Early & Burchinal, 2001; Hirshberg et. al, 2005; Greenberg, 2011; Bassok et al. 2016).[10] Differences in income and SES are likely to contribute to the gap in enrollment between Hispanic and non-Hispanic children. Hispanic families are more disadvantaged across a number of dimensions compared with white families. They have much lower income, and are much more likely to live in extreme poverty compared to white families (Radey & Brewster, 2007; U.S. Census Bureau, 2015). On average, adults in Hispanic households also have lower education levels. For instance, among Hispanic families, more than 40% of mothers do not have a high

[10] The relationship between income and preschool enrollment is not strictly linear. In particular, targeted preschool programs and subsidies have substantially altered this relationship, so that the poorest children attend preschool at *higher* rates than those with slightly more income, resulting in a U-shaped curve (e.g. Early & Burchinal, 2001).

school diploma (compared with 6% of white mothers), and only 30% had any education beyond high school (Gándara & Contreras, 2010).

While income and SES play a salient role in parents' child care decisions, it is striking that black families—whose income and home resources often mirror those of Hispanic families—actually attend center-based care at rates similar to white children (Radey & Brewster, 2007; Bassok, 2010). For instance, in 2014 39% of black children 3-5 were enrolled in preschool, compared with 41% of white children, and just 32% of Hispanic children (Kena et al, 2016). Although there are multiple hypotheses as to why black and Hispanic children differ on this dimension, the pattern suggests that factors other than family resources are likely to contribute in important ways to early enrollment patterns.

Household composition. The composition of households is another important factor that affects parents' decisions about child care. From a constraints and preferences perspective, additional adults in the household can serve as care providers, reducing the need for non-parental care in general, and by extension, the need for preschool care. For instance, households with two parents are considerably less likely to use NPC than single parent households (Liang et al, 2000; Hirshberg, Huang, & Fuller, 2005; Greenberg, 2011). Research has also shown that NPC use is also lower among households with grandparents or other non-parent relatives (Fuller et al., 1996; Liang et al., 2000; Vandell et al., 2003; Radey & Brewster, 2007).

Both the number and ages of children in each household may affect child care decisions. For instance, additional young children in the household could lead parents to decide not to work, rather than pay the additional cost of having multiple

children in NPC. It could also lead parents to place children in relative or friend care, as opposed to center-based care, to save money. Indeed, research has shown that households with more young children (0-5) are less likely to use both NPC and center-based care specifically (Greenberg, 2011; Coley et al, 2014).

It is important to note that household composition and income/SES are not wholly distinct. For instance, the presence of a second parent or the number of children in a household is associated with the resources available. However, the aspects of household composition described above have been shown to be related to enrollment gaps over and above what can be explained by SES alone.

Differences by race/ethnicity in terms of household composition are pronounced. Hispanic households are less likely to contain two parents than white households (58% vs. 75%) but considerably more likely than black households (34%) (Child Trends Databank, 2015). Hispanic and black households are more likely to contain a grandparent or other nonparent resident (Radey & Brewster, 2007), and among households that contain grandparents, Hispanic and black grandparents are more likely to provide child care (Luo, Lapierre, Hughes, & Waite, 2012). In addition, Hispanic families include more children than white families, and on average, Hispanic mothers give birth at a younger age. For example, in 2013 the birth rate among Hispanic 15-19 year olds was more than twice that for non-Hispanic whites (Martin et al., 2015).

Parent work schedules. Parents who work more need more child care. From a constraints and preferences perspective, the type of care that parents can select is partially constrained by their work schedule. For instance, parents that

work many late nights and weekends are not likely to use as much center-based care, which is available primarily during standard 9-5 hours (Hirshberg et al., 2005; Greenberg 2011).

Hispanic households differ from non-Hispanic households in terms of their employment status and work schedules. While Hispanic and white single mothers are equally likely to be employed (76%). Among married mothers, 70% of white women were in the labor force compared with 62% of Hispanic women (U.S. Bureau of Labor Statistics, 2015). Among mothers that work, white and Hispanic mothers work a similar number of hours, and a similar percentage are employed full time (Landivar 2016). However, Hispanic workers in general are disproportionately represented in the types of occupations that require irregular hours (Padilla et al., 2006).

The relationship between work outcomes and child care utilization may be bidirectional. For example, parents who feel strongly that their young children should only be home with their parents may choose to forego additional employment or income to be home with their child, or may choose non-standard work hours to facilitate their preferred care arrangement. Here, and throughout this literature, it is important to remember the potential bidirectional nature of this relationship.

The three factors described above—income, household composition, and work hours—all relate directly to families supports and constraints. Below I discuss a number of other factors that may also contribute to differential care use across groups.

Parental beliefs/preferences. Some existing research has suggested that gaps in care utilization are driven in part by differences in beliefs and values across groups, rather than solely by differences in resources.

Exposure to early literacy. For example, some of the earliest work that explored the preschool enrollment gap argued that it partly reflected differences between Hispanic and non-Hispanic families in their preferences for exposure to early literacy (Fuller et al. 1994; Fuller et al, 1996). For instance, Fuller et al. (1994) found that just 29% of Hispanic families reported reading to their child every day, compared with over 50% of white families, and suggested that Hispanic families may not place as high a value on early education and literacy.

More recent research has suggested that Hispanic and non-Hispanic parents both place high value on early literacy, but that Hispanic parents also place high value on practical features of child care, more so than white parents (Kim & Fram, 2008; Shlay, 2010; NSECE Project Team, 2014). For instance, Early & Burchinal (2001) found that Hispanic families have stronger preferences for providers that can provide care for sick children, or during flexible hours (Early & Burchinal, 2001).

Beliefs about child rearing. Other research has suggested that differences in beliefs about how best to raise children may contribute to gaps in preschool enrollment (Fuller et al., 1994; Liang et al. 2000). Hashima & Amato (1994) found that Hispanic families are more family oriented and endorse warmer parenting practices than do non-Hispanic families, which may lead them to avoid structured

settings for young children. Fuller & Garcia Coll (2010) also find differences in beliefs about parenting between Hispanic and non-Hispanic families.

Aspects of the child care market. Finally, research has speculated that in addition to differences in the demand for early care, an important driver of preschool enrollment gaps may be differences in the *availability* of care options (Fuller, Kim & Bridges, 2010). In particular, evidence suggests that Hispanic children live disproportionately in communities with lower access to early care (Bassok, Fitzpatrick, & Loeb, 2011), and that differences in access to care partially explain gaps in preschool enrollment (Hirshberg, Huang, & Fuller, 2005). At the same time, it is unclear whether lower availability of care in Hispanic communities may partially reflect lower demand among Hispanic parents, and no studies have been able to isolate the unique impact of supply.

Child's age. Parents differ in their preferences for care use among infants/toddlers (0-2) and preschool aged children (3-5, not in K). Specifically, preschoolers are considerably more likely to be enrolled in any type of NPC, and in center-based care specifically (Early et. al, 2001; Coley et al, 2014). However, parents may differ in their preferences for children of different ages by race/ethnicity. Some work, for instance, has found that cultural norms and preferences among Hispanic and immigrant households were more salient when making decisions about young children (Coley et al, 2014).

Differences among Hispanic children.

Hispanic children are quite diverse in terms of home language, immigrant status, and country of origin. For this reason, exploring differences in preschool patterns

among Hispanic children may provide insights that complement those from analyses comparing *all* Hispanic children to white or black children (Logan & Turner, 2013; Turner et al., 2015).

Home language. Hispanic families that only speak Spanish are less likely than English speakers to use both NPC and center-based care (Liang et al, 2000; Hirshberg, Huang, & Fuller, 2005). This may be at least partly because non-English speaking households face additional barriers when enrolling children in child care. For instance, Shaul (2006) found that lack of information about available child care subsidies/programs was particularly pronounced among families with limited English proficiency. A survey of low-income families in Minnesota found that households that did not primarily speak English were more likely to report having to "take whatever arrangement they could get" (Chase & Valerose, 2009).

Immigrant status. Differences across Hispanic households by immigrant status generally parallel differences by home language. For instance, nearly a third of children of immigrants live in linguistically isolated households where all members older than 13 do not speak English (Capps et al., 2004). Research has found that immigrants generally enroll their children in preschool at lower rates than non-immigrants (Brandon, 2004). Some evidence suggests that immigrant families are particularly likely to suffer from poor information about availability and quality of child care (Chaudry et al, 2011). Research from outside education has also suggested that Hispanic families face higher transaction costs when applying for public programs. For instance, Aizer (2000) found lower take up of Medicaid among Hispanic families consistent with language barriers and immigration concerns.

There are also barriers to preschool enrollment that are specific to immigrants. For instance, these families may not be eligible for the same subsidies or other opportunities as native born families, or among undocumented immigrants, they may have fears that enrolling in child care could expose their status (Karoly & Gonzalez, 2011).

The present study

Although previous research has explored the family characteristics associated with preschool enrollment gaps between Hispanic and non-Hispanic children, few studies have examined these gaps among specific Hispanic subgroups, and none have done this systematically for children of different age groups. Also, in light of widespread and dramatic changes to both the Hispanic population and the early childhood landscape since overall gaps in preschool enrollment between Hispanic and white children were examined, there is need to provide updated estimates of overall differences in preschool use. Using a new and unusually detailed nationally representative dataset of households with young children, this paper provides an updated portrait of the child care use of Hispanic, white, and black children in the United States.

Method

Data

The data for this study come from the National Survey of Early Care and Education (NSECE), a nationally representative sample of households with children under 13. This dataset includes information about 21,655 children living in 11,629 households, and allows for a detailed portrait of early child care across both home

based and center-based settings, for all 50 states and Washington, DC (NSECE Project Team 2013).[11] The sampling frame of households was based on a list of addresses maintained by the U.S. Postal Service. This list was supplemented with newly listed housing units in areas where it did not provide adequate coverage. Households were then screened for inclusion in the sampling frame with a questionnaire that asked whether there was a child under 13 in residence (NSECE Project Team, 2013).

To conduct the survey, data collectors asked to speak with an adult who was knowledgeable about the early child care usage and schedule of the youngest child in the household, with preference for a parent. The vast majority of respondents were biological mothers (approx. 95%). Respondents provided demographic information for all members of the household, including age, race, gender, and relationship to both the respondent and to children in the household.

More detailed information was collected about children under 13 and about adults who were either a parent to one of the children or a parent's spouse. In particular, the NSECE data collection included a unique feature called a "calendar" survey, which documented how these household members spent their time in 15 minute increments during an entire "reference week" (Monday-Sunday, generally

[11] The nationally representative data were collected using a multistage probability design. In the first stage, 219 primary sampling units (PSUs) were selected based on the population of children under 18. Each PSU represents either a county or group of contiguous counties. In the second stage, secondary sampling units (SSUs) were selected from within each PSU. A total of 755 SSUs were selected, and each represents either a census tract or group of contiguous tracts. SSUs were selected disproportionately from areas in which at least 40% of households had income below 250% of the federal poverty line.

the week prior to the survey). Interviews were intentionally scheduled to avoid unusual reference weeks, such as holidays.

During the calendar survey, respondents reported every person or organization who provided care for each child in the household, up to 15 providers total. For each provider-child pairing, they reported the start and stop times that care was provided during the reference week. Each child's schedule was then coded in 15 minute increments throughout the week, providing a complete picture of how children spent their time. Analogous information was collected about adults in the household. In particular, respondents reported when household members were working, in school, or in training. The calendar was administered separately for children and adults, and results were compared to resolve discrepancies. Respondents were asked to provide additional information or clarification in cases where no parents were available to provide care but a child was not with a care provider.

Measures

Race/Hispanic origin. Information about race and Hispanic origin was collected separately for all children in the sample. In keeping with previous work (e.g. Fuller et al., 1996; Early & Burchinal, 2001; Radey & Brewster, 2007). I classify all children of Hispanic origin as Hispanic, regardless of race. I refer to non-Hispanic white and non-Hispanic black children as "white" and "black" children throughout. I also consider differences within Hispanic children by home language (English only, both English & Spanish, Spanish only) and immigrant status (immigrant parent, no immigrant parent).

Child care use. I use the child calendar data described above to construct multiple measures of non-parental care use. First, I consider whether a child attended *any* non-parental care during the reference week. Next, to get a broad picture of the types of child care that are used, I consider whether a child attended any of the following: 1) center-based care, including any Head Start, public pre-k, or community-based care; 2) individual care, without a prior relationship; 3) individual care, with a prior relationship (including family/friends); 4) irregular care (<5 hours/week). Next in keeping with prior work, I consider the type of care in which children spent their most hours (the "primary" provider). Specifically, I document the proportion of children whose primary provider fell into categories 1–4 above.

Income/SES. The NSECE provides unusually rich data about the income and resources families have available. I consider some commonly used measures such as annual income, and the number of earners in each household (e.g. Fuller et al.,1996; Early & Burchinal, 2001). I also consider other measures including income to poverty ratio, the number of earners in each household, ownership of a home or car, and the proportion that received some type of public assistance [defined as food stamps, welfare, free/reduced price lunch, or WIC (Women, infants, and children) benefits]. Finally, I consider the proportion who reported not always having enough food, and the proportion that live in a community with more than 20% of families below an income to needs ratio of 1 ("high density of poverty").

Household composition. I construct a number of measures of household composition that have been found to be related to enrollment in non-parental care.

First, I consider the total number of household members, as well as the number 18 or older. I consider whether each child lived with two parents, and whether she lived with a grandparent, or other adult relative. I also consider whether children were born to teenage mothers. Here, I construct a measure of mother's age at birth by subtracting each child's current age from their mother's current age. Finally, I consider whether each child has siblings, and how many.

Parent work schedules. I use the calendar data to construct multiple measures of parent work schedules. First, to get a sense of whether any parents in the home are likely to be available to provide child care, I construct three measures that describe the combined work schedules of parents,. In these variables, a household with a single parent uses only information from that parent, whereas a household with two parents uses information from both. I consider the proportion of children who live in households where all parents work at least 40 hours/week, all parents work at least 20 hours/week, and no parents work any hours.

Next, to more closely align with previous work, I separately consider the work schedules of mothers and fathers. In particular, previous work has generally looked only at maternal employment (e.g. Fuller et al., 2000; Radey & Brewster 2007; Greenberg 2011), but I also consider paternal employment in households where a father is present.[12] Specifically, I consider the total number of hours spent working by each parent, along with an indicator for whether each parent spent *any* time working. I also consider the proportion of time each parent spent working

[12] For this analysis, I refer to respondents as "mothers" and second parents as "fathers" though that is not quite true in the data. In some cases respondents were fathers or were members of a same sex couple. However, in 95% of cases respondents were biological or step mothers.

during nonstandard hours (standard hours: 8AM-6PM Monday-Friday), overnights (10PM-6AM Monday-Sunday), and weekends.

Additional covariates. In the regression analysis, all models control for children's age in months, as children are increasingly likely to be enrolled in nonparental care as they get older, even by a few months (Liang, Fuller, & Singer, 2000). I also control for region of the country (Northeast, Midwest, West, South), and whether the community is in a rural, suburban, or urban location, to account for differences driven by household location.

Plan for analysis

Sample. Of the 21,655 children under 13 in the NSECE data, I limit my sample to children age 0-5 who had not yet started kindergarten (N= 8,939). I also limit my sample to children who had complete calendar data (approximately 97.5% of the sample), resulting in a final sample of N=8718. This sample includes 4,565 0-2 year olds and 4,153 3-5 year olds.

Descriptive comparisons. The first set of results in this paper provides nationally representative descriptive statistics comparing Hispanic children to white and black children on measures of (1) child care use, (2) income/SES (3) household composition and (4) parent work schedules. For each, I assess whether differences across groups are statistically significant. Next, I show analogous descriptive results that show differences *within* Hispanic children. Here, I separate Hispanic children by home language (English only, both English and Spanish, Spanish only) and immigrant status (lives with an immigrant parent/no immigrant parent) and test whether Hispanic children differ significantly across either

dimension. Importantly, these descriptive results do not represent the *effect* of being in an English vs. Spanish speaking household or an immigrant vs. non immigrant household. Rather, they show how households across these groups compare, allowing that the causes of these differences cannot be identified here.

Regression analysis. Next, I explore the extent to which the observed gaps in child care use (both any NPC use and in center-based care in particular) are related to differences in items (2), (3), and (4) above. Specifically, I estimate logistic regressions to predict spending the most number of hours in a center-based provider. I do this iteratively, including one set of variables at a time. I then compare how the gap in preschool enrollment differs across models to determine the extent to which each class of variables is related to this gap.

Finally, I include all four sets of explanatory variables in the same model, to examine how the gaps are related to the combination of these variables. These models take the following form:

$$\log\left(\frac{Y_i}{1-Y_i}\right) = \beta_0 + \beta_1 Hispanic_i^{'} + \beta_2 NonHisp^{'} + \beta_3 Ch_{chars_i} + \beta_4 SES_i + \beta_5 HH_{comp_i}^{'} + \beta_6 Work_i^{'} + \varepsilon_i$$

Here, I predict the probability that child *i* spends the most number of hours in center-based care. *Hispanic* is an indicator for whether a child is Hispanic, and *NonHisp* is a vector for whether a child is black, Asian, or another race. The reference group is white children. *Ch_chars'* is a vector of child characteristics including age, gender, and region of the country. *SES'* is a vector that includes income to poverty ratio, the number of earners in each household, whether a family owns a home or car, whether the family always has enough food, receipt of public

assistance, and an indicator for living in a high poverty region. *HH_comp'* includes whether a household includes two parents, a grandparent, another nonparent relative, as well as the number of siblings in the household, and an indicator for siblings 0-2 or 3-5, not in K. Finally, *Work'* contains indicators for whether all parents work 40 hours or more, 20 hours or more, and no parents work any hours, along with the number of hours that mothers and fathers spent working overall, and during nonstandard, overnight, and weekend shifts. The main coefficients of interest are the β_1s, which represent the odds that Hispanic children are enrolled in center-based care relative to white children, controlling for an extensive set of covariates.

Next I conduct analogous regressions, in which I replace the indicator *Hispanic* with multiple indicators to explore differences by home language and immigrant status. First, I estimate these models replacing *Hispanic* with indicators for whether a child's home language was English only, both English and Spanish, and Spanish only. Next, I include two indicators for whether or not a child lived with a parent who immigrated to the United States. These provide estimates of how patterns of child care use differ among Hispanic children. Again, these estimates should not be interpreted as the effect of belonging to a certain group, but rather, the difference between groups, irrespective of cause. In both set of models, the reference group remains white children.

Results

Differences by race/ethnicity.

Child care use. Table 3.1 shows descriptive statistics comparing child care use by race/ethnicity for children 0-2 and children 3-5

***Children 0-2*.** Hispanic infants and toddlers are much less likely than white or black children to attend any type of non-parental care. For instance, 44% of Hispanic children spent time in NPC, compared with 55% of black children, and 60% of white children. Hispanic children are also 7 percentage points less likely to attend center-based care than are white children (8% to 15%). When considering primary providers, Hispanic children are more likely to use an individual care provider that has a previous relationship to the household compared to white children, but *less* likely to use care from a provider with no prior relationship.

Children 3-5, not in K. There are pronounced differences in non parental care use between 0-2 year olds and preschool age children (3-5). In particular, older children are more likely to use any form of non parental care (71% to 55%), and more than three times as likely to use any type of center-based care (43% to 13%).

Patterns across race/ethnicity in preschool children are similar to those among younger kids. Hispanic children of preschool age are the least likely to attend any type of NPC (63%, compared with 74% and 76% among white and black children, respectively). They are also 9 percentage points less likely to attend center-based care than white children (46% to 37%), which is consistent with other recent estimates of this gap (Chernoff et al, 2007; Kena et al, 2016).

Income/SES. The first panel of Table 3.2 shows how household income and resources differ by race/ethnicity. Along every dimension considered, black and Hispanic children live in households that are significantly more impoverished than their white counterparts. For example, the annual income for black and Hispanic families was roughly *half* that of white families. Compared to both white and black

children, white children lived in households with more earners, that were more likely to own a house or a car, more likely to always have enough food, less likely to receive public assistance, and less likely to live in a community with a high poverty concentration.

Along many dimensions, black children live in households that are even more impoverished than Hispanic children. For example, black children live in households with lower incomes, that are less likely to own a home or care, and more likely to receive some type of public assistance.

Household composition. Table 3.2 also shows that there are substantial differences by race/ethnicity in terms of household composition. Compared to Hispanic and black children, white children are considerably more likely to live in a household with two parents (76%, 53%, and 34% for white, Hispanic, and black children, respectively). White children live in households with fewer siblings and fewer people overall. White children are also less than half as likely as black or Hispanic children to have been born to a teenage mother.

At the same time, black and Hispanic households are more likely to include grandparents or other nonparent relatives. Notably, although black children are more likely than white children to live in households with grandparents or other relatives, they still live in households with the fewest adults overall due to high levels of single-parenting. In contrast, Hispanic children lived in households with the most members overall, including the most members over 18.

Work schedules. The final panel of Table 3.2 presents information about parent work schedules. Hispanic children are more than twice as likely as white

children to live in households where no parents worked at all. Among mothers, Hispanic mothers are the least likely to work any hours. For mothers who worked, there are no differences in the average number of hours worked by race/ethnicity, but black and Hispanic mothers worked a higher proportion of overnight and weekend hours.

Differences in work schedules by race/ethnicity are considerably more pronounced among fathers. As mentioned above, Hispanic and black children are much less likely to live with two parents than are white children. This is compounded by the fact that, conditional on living with two parents, black and Hispanic fathers are also considerably less likely to work than white fathers. For example, 82% of white fathers spent any time working compared to 73% of Hispanic fathers and just 64% of black fathers. Combined with differences in likelihood of single parent households, on average just 22% of black 0-5 year olds live in households with working fathers, compared with 39% of Hispanic children and 62% of white children.[13]

Regression analysis

Table 3.3 presents results from logit estimations that model gaps in child care use across different subgroups of children. In all models, estimates are presented as odds ratios, and coefficients can be interpreted as the odds of attending a specific type of care relative to white children. In this table, I show how each category of variables (i.e. income/SES, household composition, parent work schedules) is

[13] Results not shown. These percentages are calculated as "proportion of households with two parents X proportion of fathers in two parent households who spent any hours working."

related to gaps in center-based care use between white, Hispanic, and black children. I present unadjusted gaps in model 1, and introduce a set of baseline controls in model 2. Next, I introduce each set of variables one at a time in models 3-5, finally including them all at once in model 6.

Children 0-2. Panel A demonstrates gaps in preschool use among children 0-2. Model 1 shows the unadjusted odds for Hispanic children of attending center-based care as a primary provider are 50% as high as those for white children. This difference is no longer significant after controlling for income/SES (model 3). Including controls for household composition (model 4) and parent work schedules (model 5) both explain a portion of the gap as well, though the gap remains significant when looking at these factors independently. There is no longer a significant difference in preschool use when including all sets of explanatory variables together, suggesting that together these factors account for most if not all of the differences in center-based care use.

When comparing black children 0-2 to white children, the patterns are quite different. In particular, black children are equally likely to enroll in preschool as white children even without controlling for any aspects of the household. This relationship holds as different sets of variables are introduced.

Children 3-5, not in K. Table 3.3, Panel B shows analogous results among preschool aged children. The patterns are broadly similar to those observed among younger children, but there are some differences worth noting. Hispanic children of preschool age are less likely to use both center-based care, but the gap in center-based care use is smaller among older children. Specifically, the unadjusted odds of

Hispanic preschoolers attending center-based care are 74% as high as for white children, compared with 50% among 0-2 year olds. Interestingly, there is no longer a significant difference in center-based care use when including *any* of the three sets of explanatory variables, suggesting that all three sets of explanatory variables contribute to differences in preschool enrollment.

When considering how black children 3-5 compare to white children, the results are strikingly different than for Hispanic children. For instance, without controlling for anything, black children are again enrolled in preschool at similar rates to white children. However, after controlling for income, household composition, or parent work schedules, black children are significantly *more* likely to be enrolled in preschool than are white children.

Differences among Hispanic children

Child care use. Tables 3.4 and 3.5 show how child care use varies among Hispanic children both in terms of home language and immigrant status. These measures are highly correlated. For example, 56% of immigrant households speak only Spanish, and nearly all homes that speak only Spanish are immigrant households (88%).[14] Nevertheless, disaggregating results across both groups highlights several unique patterns.

Children 0-2. Table 3.4 presents results for Hispanic children 0-2. Children from English-speaking households are considerably less likely to use any NPC than those from Spanish-speaking households (49% to 38%). They are almost twice as

[14] The first three rows of Appendix A show the extent of overlap between the two groups.

likely to use center-based care, but this difference is not significant. Patterns are quite similar by immigrant status.

Children 3-5, not in K. Table 3.5 shows differences in child care use among Hispanic preschoolers. Here, the differences by home language are striking. In particular, Hispanic children from English speaking homes are 19 percentage points more likely to use NPC than are children from Spanish speaking households (72% to 53%). They are also 14 percentage points more likely to use center-based care (44% to 30%). Further, Hispanic children from English speaking households attend both NPC and center-based care at rates very similar to those of white children (Table 3.1). This shows that the gap in preschool enrollment is mostly due to Hispanic households that are dual language or Spanish only. However, this should not be interpreted to mean that being in a dual language or Spanish household *causes* lower enrollment in center based care.

By contrast, there are no significant differences between children from immigrant an non-immigrant households, which is itself striking given the strong differences seen by home language and the overlap between home language and immigration status.

Family characteristics, by Hispanic subgroups. Appendix 3A presents differences in other family characteristics by home language and immigration status. As expected, many patterns for Hispanic children in Spanish-only/dual language (hereafter "Spanish speaking") households and immigrant households are similar. Both Spanish speaking and immigrant households were significantly more impoverished across many of the income/SES measures, relative to English

speaking and non-immigrant households, respectively. These households were also larger and more likely to include non-parent relatives.

The bottom panel of Appendix 3A shows differences in parent work schedules by Hispanic subgroup. For both Spanish speaking and immigrant households, mothers are less likely to work. However, Spanish speaking households are also more likely than English speaking households to have NO parents working. By comparison, there is not a significant difference between children from immigrant and non-immigrant households.

Regression analysis

Tables 3.6 & 3.7 present results analogous to those shown in Table 3.3 above, but separately for Hispanic subgroups. I show separate coefficients for English only, dual language, and Spanish only households, as well as for immigrant and non-immigrant households. For comparison, the top row of each table replicates the corresponding coefficients from Table 3.3. Coefficients are still interpreted as odds of attending a certain type of care relative to white children.

Children 0-2. Table 3.6 shows results among Hispanic children 0-2. There are striking differences by subgroup. English-only households are considerably more likely to use center-based care as a primary provider than are Spanish-only or dual language households. Patterns are similar by immigration status.

In looking across models 2-5, income and SES is the factor that is most related to gaps in preschool enrollment. For instance, in model 3, which controls only for income/SES there are no longer any significant differences in center-based care for any home language subgroup. By contrast, after controlling for income and

SES, Hispanic children from immigrant households are still somewhat less likely to attend center-based care than white children, though the difference is no longer significant.

Children 3-5, not in K. Finally, Table 3.7 shows differences among Hispanic preschoolers. There are pronounced differences in terms of home language. For instance, children from English-only households use center-based care at rates roughly equal to white children, but the odds of children from Spanish-only households enrolling in center-based care are only 60% as large as the odds for white children. By contrast, there are no noteworthy differences in care use by immigrant status.

In examining which sets of covariates are related to gaps between Hispanic subgroups and white children, income and SES again are again most strongly associated with these gaps. Strikingly, once all sets of covariates are controlled for, all sets of Hispanic subgroups attend center-based care at rates almost identical to those of white children.

Discussion & Policy Implications

This project uses new, nationally representative data to document differences in child care use across race/ethnicity and across different Hispanic subgroups. I document these patterns separately for infants/toddlers and preschool aged children. Next, I examine the relationship between family characteristics and gaps in preschool enrollment separately among Hispanic subgroups.

I find that Hispanic children of all ages are less likely than white children to attend preschool. As expected, there are notable differences in child care use across

different types of Hispanic households. In particular, immigrant and Spanish-speaking households are the least likely to enroll children in preschool. By comparison, Hispanic children from English-speaking households enroll at roughly the same rates as white children. Along other dimensions of child care use and along measures of income and SES as well, Hispanic children from English-speaking households are more similar to white children than to children from Spanish-speaking households. By contrast, black children of all ages are equally or more likely than white children to be enrolled in preschool care.

As past research has found, income and socioeconomic status are strongly related to gaps in preschool enrollment (Radey & Brewster, 2007; Greenberg, 2011). In many cases, after controlling for these variables I find no differences in enrollment rates between white and Hispanic children. These results are purely descriptive, and should not be interpreted as evidence that income plays a causal role in these enrollment gaps. To a lesser extent, I find that household composition and parent work schedules are also related to differences in preschool use.

It is interesting to note that after adjusting for family characteristics, older Hispanic children of all types attend preschool at rates identical to white children. This suggests that differences in parental beliefs and preferences across Hispanic and non-Hispanic households may not play much of a role in whether 3-5 year olds are enrolled in preschool. By contrast, Hispanic infants and toddlers, and particularly those from Spanish-speaking and immigrant households, are less likely to attend preschool than white children even after accounting for family characteristics. This suggests that other factors beyond what have been considered

here are likely contributing to this gap. In particular, there may be differences in cultural beliefs/practices about parenting, or differences in access to information or availability of child care (Fuller et al., 1996; Harwood et al., 2002; Fuller & Garcia-Coll, 2010; Coley et al, 2014).

As mentioned above, the evidence is not clear that preschool care is always beneficial for the youngest children, and some studies have found negative effects of these programs on social and emotional outcomes among infants/toddlers (e.g. Phillips & Adams, 2001; Love et al, 2003). As such, it is important to note that more preschool care is not necessarily the best option for all families, and we should be cautious in framing differences in preschool enrollment among these children as a gap that must be remedied. Indeed, young Hispanic children are likely to receive benefits from parent or relative care that are not directly captured by measures of academic performance. More research needs to be done to determine the extent to which gaps in preschool use among infants and toddlers should be viewed as problematic.

Despite the recent increases in targeted preschool for underserved populations, it is clear that children from Spanish-speaking and immigrant households attend center based care at lower rates than white children. It will be important for future research to attempt to delineate between possible explanations for differences in enrollment among these specific subgroups. In particular, are there barriers to enrollment in preschool that are specific to immigrant or non-English speaking households? Future research should consider which specific barriers these types of households face, with an eye toward informing ways in

which they can be overcome. Future work should also consider policies such as information campaigns targeting immigrant and Spanish speaking households.

The current findings provide additional support for recent calls to consider differences among Hispanic children, rather than treating them as a single group (Logan & Turner, 2013; Mendez, Crosby, & Helms, 2016). For instance, although patterns by home language and immigration status are similar, gaps in preschool enrollment among older children seem to be more pronounced when considering differences by home language than by immigrant status. These types of nuances are overlooked when considering only averages across all Hispanic children. In addition to home language and immigrant status, researchers should consider differences among Hispanics by country of origin where possible.

There are a few important limitations to this research. First, the patterns documented are descriptive. Many of the measures I use are likely to be highly correlated and so it is not possible to make any claims to causality. Further, there are likely to be other, unmeasured characteristics of families or their environments that contribute to differences in preschool enrollment. In particular, I do not examine characteristics of the child care supply or family beliefs/preferences, both of which have been shown to be related to early child care selection as well as to SES. Fortunately, the restricted-use version of the NSECE will ultimately provide information relating both to the supply of early care *and* family beliefs preferences. Although these data are not yet available, they will shed additional light on why enrollment in preschool differs by race and ethnicity.

Also, this work is limited by the cross sectional nature of the data. In particular, child care use likely differs on dimensions other than setting, including both the length and timing of care. The current data does not allow me to examine differences along these dimensions.

Despite these limitations, this paper adds to our understanding of why Hispanic children are enrolled in preschool at lower rates, and provides the most comprehensive exploration to date of preschool use among a diverse Hispanic population of infants, toddlers and preschoolers.

References

Ansari, A., & Winsler, A. (2012). School readiness among low-income, Latino children attending family childcare versus centre-based care. *Early Child Development and Care*, *182*(11), 1465–1485. http://doi.org/10.1080/03004430.2011.622755

Barnett, W. S., Carolan, M., Squires, J., Clarke Brown, K., & Horowitz, M. (2015). The state of preschool 2014. *The National Institute for Early Education Research*.

Bassok, D. (2010). Do black and Hispanic children benefit more from preschool? Understanding differences in preschool effects across racial groups. *Child Development*, *81*(6), 1828–1845. http://doi.org/10.1111/j.1467-8624.2010.01513.x

Bassok, D., Finch, J., Lee, R., Reardon, S., & Waldfogel, J. (n.d.). Socioeconomic gaps in early childhood experiences, 1998 to 2010. *EdPolicyWorks Working Paper Series*. Retrieved from http://curry.virginia.edu/uploads/resourceLibrary/38_Gaps_In_Early_Experience.pdf

Bassok, D., Fitzpatrick, M., Greenberg, E., & Loeb, S. (2016). Within- and between-sector quality differences in early childhood education and care. *Child Development*, n/a-n/a. http://doi.org/10.1111/cdev.12551

Bassok, D., Fitzpatrick, M., & Loeb, S. (2012). Disparities in child care availability across communities: Differential reflection of targeted interventions and local demand. *Unpublished Paper, University of Virginia*.

Bassok, D., & Latham, S. (2014). Kids today: Changes in school-readiness in an early childhood era. *EdPolicyWorks Working Paper Series*.

Becerra, R. M., & Chi, I. (1992). Child care preferences among low-income minority families. *International Social Work*, *35*(1), 35–47. http://doi.org/10.1177/002087289203500104

Black, D., Haviland, A., Sanders, S., & Taylor, L. (2006). Why do minority men earn less? A study of wage differentials among the highly educated. *Review of Economics and Statistics*, *88*(2), 300–313. http://doi.org/10.1162/rest.88.2.300

Bloom, H. S., & Weiland, C. (2015). Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study. *Available at SSRN 2594430*. Retrieved from http://www.mdrc.org/sites/default/files/quantifying_variation_in_head_start.pdf

Brandon, P. D. (2004). The child care arrangements of preschool-age children in immigrant families in the United States. *International Migration*, *42*(1), 65–87. http://doi.org/10.1111/j.0020-7985.2004.00274.x

Bridges, M., Fuller, B., Rumberger, R., & Tran, L. (2004). *Preschool for California's children: promising benefits, unequal access*. Policy Analysis for California Education, PACE. Retrieved from http://eric.ed.gov/?id=ED491703

Burchinal, M. R., Roberts, J. E., Riggins, J., Rhodus, Zeisel, S. A., Neebe, E., & Bryant, D. (2000). Relating Quality of Center-Based Child Care to Early Cognitive and Language Development Longitudinally. *Child Development*, *71*(2), 339–357. http://doi.org/10.1111/1467-8624.00149

Buriel, R., & Hurtado-Ortiz, M. T. (2000). Child care practices and preferences of native- and foreign-born Latina mothers and Euro-American mothers. *Hispanic Journal of Behavioral Sciences*, *22*(3), 314–331. http://doi.org/10.1177/0739986300223003

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, *112*(3), 579–620.

Capizzano, J., Adams, G., & Ost, J. (2006). Caring for children of color. Retrieved from ftp://timecard.urban.org/pubs_prod/2006/PDF/batch37/311285_OP-72.pdf

Carneiro, P., & Heckman, J. (2003). Human capital policy. *NBER Working Paper 9495*. Retrieved from http://www.nber.org/papers/w9495.pdf

Chase, R., & Valerose, J. (2010). *Child care use in Minnesota: Report of the 2009 Statewide Household Child Care Survey*. Minnesota Department of Human Services. Retrieved from https://www.wilder.org/Wilder-Research/Publications/Studies/Child%20Care%20Use%20in%20Minnesota%202009/Child%20Care%20Use%20in%20Minnesota,%20Full%20Report.pdf

Chaudry, A., Pedroza, J. M., Sandstrom, H., Danziger, A., Grosz, M., Scott, M., & Ting, S. (2011). *Child care choices of low-income working families*. Urban Institute. Retrieved from http://www.urban.org/sites/default/files/alfresco/publication-pdfs/412343-Child-Care-Choices-of-Low-Income-Working-Families.PDF

Chernoff, J. J., Flanagan, K. D., McPhee, C., & Park, J. (2007). Preschool: First findings from the preschool follow-up of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B). *National Center for Education Statistics*. Retrieved

from http://eric.ed.gov/?id=ED498793

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2010). *How does your kindergarten classroom affect your earnings? Evidence from Project STAR* (Working Paper No. 16381). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w16381

Child Trends Databank. (2014). Early childhood program enrollment. Retrieved from http://www.childtrends.org/?indicators=early-childhood-program-enrollment

Child Trends Databank. (2015). *Family structure*. Retrieved from http://www.childtrends.org/?indicators=family-structure

Coley, R. L., Votruba-Drzal, E., Collins, M. A., & Miller, P. (2014). Selection into early education and care settings: Differences by developmental period. *Early Childhood Research Quarterly*, *29*(3), 319–332. http://doi.org/10.1016/j.ecresq.2014.03.006

Davis, E. E., & Connelly, R. (2005). The influence of local price and availability on parents' choice of child care. *Population Research and Policy Review*, *24*(4), 301–334. http://doi.org/10.1007/s11113-005-8515-y

Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics*, *1*(3), 111–134. http://doi.org/10.1257/app.1.3.111

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. http://doi.org/10.1037/0012-1649.43.6.1428

Early, D. M., & Burchinal, M. R. (2001). Early childhood care: relations with family characteristics and preferred care characteristics☆. *Early Childhood Research Quarterly*, *16*(4), 475–497. http://doi.org/10.1016/S0885-2006(01)00120-X

Federal Interagency Forum on Child and Family Statistics. (2015). *America's children: Key national indicators of well-being, 2015*. Retrieved from http://www.childstats.gov/americaschildren/tables.asp

Fuller, B., Eggers-Pierola, C., Holloway, S., & Liang, X. (1994). Rich culture, poor markets: Why do Latino parents choose to forego preschooling?. Retrieved from zotero://attachment/6/

Fuller, B., & García Coll, C. (2010). Learning from Latinos: Contexts, families, and child development in motion. *Developmental Psychology*, *46*(3), 559–565. http://doi.org/10.1037/a0019412

Fuller, B., Holloway, S. D., & Liang, X. (1996). Family selection of child-care centers: the influence of household support, ethnicity, and parental practices. *Child Development*, *67*(6), 3320–3337.

Fuller, B., & Kim, A. Y. (2011). Latino access to preschool stalls after earlier gains: Certain to harden achievement gaps, erode workforce quality. New Journalism on Latino Children. *Institute of Human Development (NJ1)*. Retrieved from http://eric.ed.gov/?id=ED538872

Fuller, B., Kim, Y., & Bridges, M. (2010). Few preschool slots for Latino children: scarce access in Illinois drives learning gaps, even before starting school. New Journalism on Latino Children. *Institute of Human Development (NJ1)*. Retrieved from http://eric.ed.gov/?id=ED538870

Gándara, P., & Contreras, F. (2010). *The Latino education crisis*. Harvard University Press. Retrieved from http://www.hup.harvard.edu/catalog.php?isbn=9780674047051

Gormley, W. T., & Phillips, D. (2005). The effects of universal pre-k in Oklahoma: Research highlights and policy implications. *Policy Studies Journal*, *33*(1), 65–82.

Greenberg, J. P. (2011). The impact of maternal education on children's enrollment in early childhood education and care. *Children and Youth Services Review*, *33*(7), 1049–1057. http://doi.org/10.1016/j.childyouth.2011.01.016

Harwood, R., Leyendecker, B., Carlson, V., Asencio, M., & Miller, A. (2002). *Parenting among Latino families in the U.S.* Psychology Press.

Hashima, P. Y., & Amato, P. R. (1994). Poverty, social support, and parental behavior. *Child Development*, *65*(2), 394–403. http://doi.org/10.2307/1131391

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902. http://doi.org/10.1126/science.1128898

Hemphill, F. C., & Vanneman, A. (2011). Achievement gaps: How Hispanic and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress. Statistical analysis report. NCES 2011-459. *National Center for Education Statistics*. Retrieved from http://eric.ed.gov/?id=ED520960

Hirshberg, D., Huang, D. S.-C., & Fuller, B. (2005). Which low-income parents select child-care? *Children and Youth Services Review*, *27*(10), 1119–1148. http://doi.org/10.1016/j.childyouth.2004.12.029

Illinois Facilities Fund. (2003). *We need more day care centers: Latino families and child care preferences in metropolitan Chicago*.

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., … Velez, E. D. (2016). *The condition of education 2016*. Government Printing Office.

Kim, J., & Fram, M. S. (2009). Profiles of choice: Parents' patterns of priority in child care decision-making. *Early Childhood Research Quarterly*, *24*(1), 77–91. http://doi.org/10.1016/j.ecresq.2008.10.001

Landivar, L. C. (2013). Who opts out? Labor force participation among Asian, Black, Hispanic, and White mothers in 20 occupations. *Notions of Family: Intersectional Perspectives. Advances in Gender Research*, *17*, 263–286.

Liang, X., Fuller, B., & Singer, J. D. (2000). Ethnic differences in child care selection: The influence of family structure, parental practices, and home language. *Early Childhood Research Quarterly*, *15*(3), 357–384.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*(1), 52–66. http://doi.org/10.1016/j.econedurev.2005.11.005

Logan, J. R., & Turner, R. N. (2013). Hispanics in the United States: not only Mexicans. *US2010 Project*. Retrieved from http://www.s4.brown.edu/us2010/Data/Report/report03202013.pdf

Love, J. M., Harrison, L., Sagi-Schwartz, A., Van IJzendoorn, M. H., Ross, C., Ungerer, J. A., … others. (2003). Child care quality matters: How conclusions may vary with context. *Child Development*, *74*(4), 1021–1033.

Luo, Y., LaPierre, T. A., Hughes, M. E., & Waite, L. J. (2012). Grandparents providing care to grandchildren: A population-based study of continuity and change. *Journal of Family Issues*, *33*(9), 1143–1167. http://doi.org/10.1177/0192513X12438685

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

Magnuson, K., Lahaie, C., & Waldfogel, J. (2006). Preschool and school readiness of children of immigrants. *Social Science Quarterly*, *87*(5), 1241–1262.

http://doi.org/10.1111/j.1540-6237.2006.00426.x

Martin, J., Hamilton, B., Osterman, M., Curtin, S., & Matthews, T. J. (2015). *Births: Final data for 2013* (National Vital Statistics Reports No. 1). U.S. Department of Health and Human Services. Retrieved from http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_01.pdf

Mendez, J., Crosby, D., & Helms, H. (2016). Families' utilization of early care and education. Retrieved from http://childtrends.org/wp-content/uploads/2016/02/ECE-Series-Brief-No.-3.pdf

NSECE project team. (2013). *National Survey of Early Care and Education: Summary of data collection and sampling methodology* (OPRE Report No. 2013–46). Washington DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

NSECE project team. (2014). *Household search for and perceptions of early care and education: Initial findings from the National Survey of Early Care and Education (NSECE)* (OPRE Report No. 2014–55a). Washington DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

O'Hare, W. (2011). *The changing child population of the United States: Analysis of data from the 2010 Census*. The Annie E. Casey Foundation.

Padilla, Y. C., Radey, M. D., Hummer, R. A., & Kim, E. (2006). The living conditions of U.S.-born children of Mexican immigrants in unmarried families. *Hispanic Journal of Behavioral Sciences*, *28*(3), 331–349. http://doi.org/10.1177/0739986306290367

Phillips, D., & Adams, G. (2001). Child care and our youngest children. *The Future of Children*, 35–51.

Radey, M., & Brewster, K. L. (2007). The influence of race/ethnicity on disadvantaged mothers' child care arrangements. *Early Childhood Research Quarterly*, *22*(3), 379–393. http://doi.org/10.1016/j.ecresq.2007.05.004

Reardon, S. F., & Galindo, C. (2009). The Hispanic-white achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*(3), 853–891. http://doi.org/10.3102/0002831209333184

Reardon, S. F., & Portilla, X. (2015). Recent trends in socioeconomic and racial school readiness gaps at kindergarten entry. In *Annual Meeting of the Society for Research on Educational Effectiveness*. Retrieved from https://cepa.stanford.edu/sites/default/files/reardon%20portilla%20school%20readiness%20gap%20trends%2011sep2015.pdf

Reardon, S., Robinson-Cimpian, J., & Weathers, E. (2015). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In H. Ladd & M. Goertz (Eds.), *Handbook of research in education finance and policy, 2nd edition*.

Rooney, P., Hussar, W., Planty, M., Choy, S., Hampden-Thompson, G., Provasnik, S., & Fox, M. A. (2006). *The condition of education, 2006* (No. 2006–71). National Center for Education Statistics. Retrieved from http://eric.ed.gov/?id=ED491909

Schweinhart, L., Montie, J., Xiang, Z., Barnett, W., Belfield, C., & Nores, M. (2005). Lifetime effects: The High/Scope Perry Preschool study through age 40. *High/Scope Educational Research Foundation*.

Shaul, M. S. (2006). *Child care and early childhood education: More information sharing and program review by HHS could enhance access for families with limited English proficiency.* (No. GAO-06-807). Government Accounting Office. Retrieved from http://eric.ed.gov/?id=ED493581

Shlay, A. B. (2010). African American, White and Hispanic child care preferences: A factorial survey analysis of welfare leavers by race and ethnicity. *Social Science Research*, *39*(1), 125–141. http://doi.org/10.1016/j.ssresearch.2009.07.005

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital*, *1*(1), 91–136. http://doi.org/10.1086/526401

Turner, K., Guzman, L., Wildsmith, E., & Scott, M. (2015). The complex and varied households of low-income Hispanic children. *Washington, DC: Child Trends. Http://www. Childtrends. org/wpcontent/uploads/2015/01/2015-04ComplexHouseholdsLowIncomeHispanic. Pdf*, *2100*. Retrieved from http://www.childtrends.org/wp-content/uploads/2015/01/2015-04ComplexHouseholdsLowIncomeHispanic.pdf

U.S. Bureau of Labor Statistics. (2015, November). Labor force characteristics by race and ethnicity, 2014. Retrieved June 7, 2016, from http://www.bls.gov/opub/reports/race-and-ethnicity/archive/labor-force-characteristics-by-race-and-ethnicity-2014.pdf

U.S. Census Bureau. (2011). *Overview of race and Hispanic origin: 2010* (American Community Survey Briefs No. ACSBR/10-19). U.S. Census Bureau.

U.S. Census Bureau. (2012). *Population with bachelors degree or higher by race and Hispanic origin*. Retrieved from

http://www.census.gov/library/publications/2012/acs/acsbr10-19.html

U.S. Census Bureau. (2014). *Projections of the size and composition of the U.S. population: 2014 to 2060* (No. P25-1143). Retrieved from http://www.census.gov/library/publications/2015/demo/p25-1143.html

U.S. Census Bureau. (2015). *Income and poverty in the United States: 2014* (No. P60-252). Retrieved from http://www.census.gov/library/publications/2015/demo/p60-252.html

Vandell, D. L., McCartney, K., Owen, M. T., Booth, C., & Clarke-Stewart, A. (2003). Variations in child care by grandparents during the first three years. *Journal of Marriage and Family*, *65*(2), 375–381. http://doi.org/10.1111/j.1741-3737.2003.00375.x

Vogel, C. A., Aikens, N., Burwick, A., Hawkinson, L., Richardson, A., Mendenko, L., & Chazan-Cohen, R. (2006). *Findings from the Survey of Early Head Start Programs: Communities, Programs, and Families. Final Report*. US Department of Health and Human Services. Retrieved from http://eric.ed.gov/?id=ED498072

Watamura, S. E., Donzella, B., Alwin, J., & Gunnar, M. R. (2003). Morning-to-afternoon increases in cortisol concentrations for infants and toddlers at child care: Age differences and behavioral correlates. *Child Development*, *74*(4), 1006–1020.

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), 2112–2130. http://doi.org/10.1111/cdev.12099

West, J., Germino Hausken, E., & Collins, M. (1993). *Profile of preschool children's child care and early education program participation*. Office of Educational Research and Improvements

Table 3.1 Child care use among children 0-5, by race/ethnicity

	0-2 year olds				3-5 year olds, not in K			
Child care variables	All	White	Hispanic	Black	All	White	Hispanic	Black
Any time spent in…								
Non-parental care (NPC)	0.55 **	0.60 a	0.44 ac	0.55 c	0.71 **	0.74 a	0.63 ac	0.76 c
Center based care	0.13 **	0.15 a	0.08 ac	0.14 c	0.43 **	0.46 a	0.37 ac	0.45 c
Individual care, no prior relationship	0.08	0.10 ab	0.04 a	0.06 b	0.06	0.08 ab	0.03 a	0.04 b
Individual care, prior relationship	0.26	0.26	0.24	0.28	0.24	0.24	0.24	0.27
Irregular care (< 5 hours/week)	0.15	0.19 ab	0.09 a	0.11 b	0.15	0.19 ab	0.10 a	0.10 b
Conditional on any hours in NPC:								
Primary provider (most hours) is…								
Center based care	0.22 **	0.23	0.17	0.24	0.51 **	0.51	0.50	0.55
Individual care, no prior relationship	0.13 **	0.15 a	0.08 a	0.10	0.08 **	0.09 ab	0.04 a	0.05 b
Individual care, prior relationship	0.42 **	0.37 a	0.53 a	0.46	0.26 **	0.25	0.31	0.24
Irregular care	0.18 **	0.21 b	0.16	0.13 b	0.08 **	0.09	0.09	0.06
N	4565	1851	1534	728	4153	1720	1384	671

Note. Columns that share a superscript are significantly different from each other at p<.05. Differences across age groups are denoted as follows:
* p<.05 ** p<.01 *** p<.001

Table 3.2 Differences in family predictors of early child care use, by race/ethnicity

	All	White	Hispanic	Black
Income/SES				
Annual income in 2011 (thousands)	52.8	64.1 [ab]	36.5 [ac]	31.3 [bc]
Income to poverty ratio	2.54	3.13 [ab]	1.69 [a]	1.47 [b]
Number of earners in HH	1.52	1.58 [ab]	1.45 [a]	1.41 [b]
Family owns their own home	0.47	0.62 [ab]	0.32 [ac]	0.20 [bc]
Family owns a car	0.87	0.94 [ab]	0.80 [ac]	0.71 [bc]
Family does not always have enough food	0.11	0.07 [ab]	0.17 [a]	0.17 [b]
Family receives public assistance	0.50	0.34 [ab]	0.71 [ac]	0.79 [bc]
High density of poverty in community	0.22	0.11 [ab]	0.34 [ac]	0.47 [bc]
Household composition				
Total household members	4.57	4.39 [ab]	4.94 [ac]	4.61 [bc]
Total household members over 18	2.14	2.10 [ab]	2.35 [ac]	1.94 [bc]
Child lives w/two parents	0.65	0.76 [ab]	0.53 [ac]	0.34 [bc]
Parent 1 was < 20 at child's birth	0.08	0.05 [ab]	0.12 [a]	0.14 [b]
Household includes grandparent of child	0.09	0.07 [ab]	0.12 [a]	0.16 [b]
Household includes other adult relatives (e.g aunts/cousins)	0.19	0.13 [ab]	0.31 [ac]	0.21 [bc]
Any siblings in household	0.74	0.73 [a]	0.77 [ac]	0.72 [c]
Number of siblings in household	1.25	1.16 [ab]	1.36 [a]	1.39 [b]
N	8718	3571	2918	1399

Note. Public assistance is defined as receipt of food stamps, welfare, free/reduced price lunch, or WIC (women, infant, children) benefits. High density of poverty defined as a community where >20% of families have income to needs ratio below 1. Columns that share a superscript are significantly different from each other at $p < .05$

Table 3.2 Differences in family predictors of early child care use, by race/ethnicity (continued)

	All	White	Hispanic	Black
Parent work schedules				
All parents in HH worked at least 40 hrs/wk	0.21	0.21	0.18 [c]	0.24 [c]
All parents in HH worked at least 20 hrs/wk	0.35	0.36 [a]	0.31 [ac]	0.38 [c]
No parents in HH worked any hours	0.23	0.15 [ab]	0.34 [a]	0.39 [b]
Mother's work schedule				
Any hours worked	0.53	0.55 [a]	0.46 [a]	0.52
Conditional on any hours worked:				
Total hours worked	37.59	36.93	38.63	37.94
Proportion of work during nonstandard hrs+	0.24	0.21 [ab]	0.27 [a]	0.30 [b]
Proportion of work during overnight hrs+	0.04	0.04 [b]	0.04 [c]	0.08 [bc]
Proportion of work during weekend hrs	0.10	0.08 [a]	0.12 [a]	0.10
Father's work schedule				
Conditional on father in household:				
Any hours worked	0.77	0.82 [ab]	0.73 [a]	0.64 [b]
Conditional on any hours worked:				
Total hours worked	46.63	46.84	45.25	49.04
Proportion of work during nonstandard hrs	0.24	0.23 [b]	0.25	0.29 [b]
Proportion of work during overnight hrs	0.05	0.04 [b]	0.05 [c]	0.08 [bc]
Proportion of work during weekend hrs	0.08	0.08	0.08	0.10
N	8718	3571	2918	1399

Note. + Standard hours: 8AM-6PM Mon-Fri, Overnight hours: 10PM-6AM Mon-Sun.
Columns that share a superscript are significantly different from each other at p<.05

Table 3.3 Odds of Hispanic and black children 0-5 attending primarily center-based care, relative to white children

Panel A. Children 0-2

	Primary care provider is center-based					
	(1)	(2)	(3)	(4)	(5)	(6)
Hispanic	0.50***	0.49***	0.81	0.64*	0.65*	0.86
	(0.09)	(0.10)	(0.16)	(0.13)	(0.14)	(0.18)
Black	0.94	0.75	1.49	0.79	1.10	1.05
	(0.20)	(0.16)	(0.36)	(0.19)	(0.25)	(0.27)
N	4565	4565	4565	4565	4565	4565

Panel B. Children 3-5, not in K

	Primary care provider is center-based					
	(1)	(2)	(3)	(4)	(5)	(6)
Hispanic	0.74**	0.67**	0.96	0.87	0.83	1.03
	(0.09)	(0.09)	(0.13)	(0.12)	(0.11)	(0.15)
Black	1.14	1.01	1.58**	1.38*	1.41*	1.69**
	(0.16)	(0.15)	(0.26)	(0.21)	(0.22)	(0.28)
N	4153	4153	4153	4153	4153	4153
Age, gender, region, urbanicity		X	X	X	X	X
Income/SES measures			X			X
Household composition				X		X
Parent work schedules					X	X

Note. Results come from logit estimates of the likelihood of attending center-based care. All models include indicators for whether a child was Hispanic, black, Asian, or some other race (nonwhite), so coefficients can be interpreted as odds relative to white children. * p<.05 ** p<.01 *** p<.001

Table 3.4 Child care use among Hispanic children 0-2, by home language and immigrant status

	Home language			Immigrant parent	
Child care variables	English only	English & Spanish	Spanish only	No	Yes
Any time spent in…					
Non-parental care (NPC)	0.49 [b]	0.41	0.38 [b]	0.52 [d]	0.36 [d]
Center-based care	0.11	0.07	0.06	0.11	0.06
Individual care, no prior relationship	0.04	0.04	0.03	0.04	0.03
Individual care, prior relationship	0.27	0.26	0.21	0.29 [d]	0.20 [d]
Irregular care (< 5 hours/week)	0.14 [ab]	0.06 [a]	0.06 [b]	0.13 [d]	0.05 [d]
Conditional on any hours in NPC:					
Primary provider (most hours) is…					
Center-based care	0.19	0.16	0.13	0.18	0.15
Individual care, no prior relationship	0.09	0.06	0.09	0.08	0.09
Individual care, prior relationship	0.49	0.59	0.53	0.51	0.55
Irregular care	0.19	0.12	0.15	0.18	0.13
N	498	412	608	658	850

Note. Columns that share a superscript are significantly different from each other at p<.05. Superscripts *a,b,c* indicate differences by home language and *d* indicates differences by immigrant status.

Table 3.5 Child care use among Hispanic children 3-5, by home language and immigrant status

	Home language			Immigrant parent	
Child care variables	English only	English & Spanish	Spanish only	No	Yes
Any time spent in…					
Non-parental care (NPC)	0.72 [ab]	0.61 [a]	0.53 [b]	0.66	0.60
Center-based care	0.44 [ab]	0.33 [a]	0.30 [b]	0.38	0.35
Individual care, no prior relationship	0.05 [a]	0.01 [a]	0.02	0.03	0.03
Individual care, prior relationship	0.28 [b]	0.23	0.18 [b]	0.28	0.20
Irregular care (< 5 hours/week)	0.14 [b]	0.10	0.06 [b]	0.13	0.08
Conditional on any hours in NPC:					
Primary provider (most hours) is…					
Center-based care	0.50	0.48	0.52	0.47	0.52
Individual care, no prior relationship	0.04	0.01	0.04	0.03	0.04
Individual care, prior relationship	0.31	0.32	0.29	0.32	0.30
Irregular care	0.08	0.12	0.08	0.10	0.08
N	462	374	532	583	774

Note. Columns that share a superscript are significantly different from each other at p<.05. Superscripts *a,b,c* indicate differences by home language and *d* indicates differences by immigrant status.

Table 3.6 Odds of Hispanic children primarily attending center based care relative to white children, among children 0-2

	Primary care provider is center-based					
	(1)	(2)	(3)	(4)	(5)	(6)
All Hispanic children	0.50***	0.49***	0.81	0.64*	0.65*	0.86
	(0.09)	(0.10)	(0.16)	(0.13)	(0.14)	(0.18)
Home language						
English only	0.63	0.69	0.84	0.80	0.80	0.94
	(0.17)	(0.20)	(0.22)	(0.22)	(0.23)	(0.25)
English & Spanish	0.43**	0.40**	0.68	0.53*	0.57	0.78
	(0.12)	(0.11)	(0.20)	(0.15)	(0.17)	(0.25)
Spanish only	0.33***	0.32***	0.72	0.45**	0.45**	0.64
	(0.09)	(0.09)	(0.21)	(0.12)	(0.14)	(0.20)
Immigrant status						
Non immigrant household	0.66	0.69	0.95	0.84	0.8	0.96
	(0.16)	(0.17)	(0.22)	(0.21)	(0.20)	(0.24)
Immigrant household	0.35***	0.33***	0.66	0.45***	0.49**	0.68
	(0.08)	(0.08)	(0.17)	(0.11)	(0.13)	(0.19)
N	4565	4565	4565	4565	4565	4565
Age, gender, region, urbanicity		X	X	X	X	X
Income/SES			X			X
Household composition				X		X
Parent work schedules					X	X

Note. Results come from logit estimates of the likelihood of attending each type of care. All models include indicators for whether a child was Hispanic, black, Asian, or some other race (nonwhite), so coefficients can be interpreted as odds relative to white children. Immigrant household refers to children who live with at least one parent who immigrated to the United States.
* p<.05 ** p<.01 *** p<.001

Table 3.7 Odds of Hispanic children primarily attending center-based care relative to white children, among children 3-5, not yet in K

	Primary care provider is center-based					
	(1)	(2)	(3)	(4)	(5)	(6)
All Hispanic children	0.74**	0.67**	0.96	0.87	0.83	1.03
	(0.09)	(0.09)	(0.13)	(0.12)	(0.11)	(0.15)
Home language						
English only	0.91	0.85	1.08	1.05	0.97	1.15
	(0.15)	(0.15)	(0.20)	(0.19)	(0.18)	(0.21)
English & Spanish	0.65*	0.58**	0.81	0.69	0.72	0.86
	(0.13)	(0.12)	(0.17)	(0.14)	(0.16)	(0.19)
Spanish only	0.60**	0.55***	0.91	0.82	0.74	1.00
	(0.10)	(0.09)	(0.17)	(0.15)	(0.14)	(0.19)
Immigrant status						
Non immigrant household	0.73*	0.67*	0.89	0.87	0.78	0.97
	(0.11)	(0.11)	(0.15)	(0.15)	(0.13)	(0.16)
Immigrant household	0.73*	0.66**	1.00	0.86	0.85	1.07
	(0.11)	(0.10)	(0.17)	(0.14)	(0.14)	(0.19)
N	4153	4153	4153	4153	4153	4153
Age, gender, region, urbanicity		X	X	X	X	X
Income/SES			X			X
Household composition				X		X
Parent work schedules					X	X

Note. Results come from logit estimates of the likelihood of attending each type of care. All models include indicators for whether a child was Hispanic, black, Asian, or some other race (nonwhite), so coefficients can be interpreted as odds relative to white children. Immigrant household refers to children who live with at least one parent who immigrated to the United States.
* p<.05 ** p<.01 *** p<.001

Appendix 3A. Characteristics of Hispanic households by home language and immigrant status

	Home language			Immigrant parent	
Variable	English only	English & Spanish	Spanish only	No	Yes
Overlap between home language/immigration					
Child lives w/parent who immigrated to U.S.	0.14 [ab]	0.61 [ac]	0.88 [bc]	0.00	1.00
Both English & Spanish are spoken in home	0.00	1.00	0.00	0.23 [d]	0.33 [d]
Only Spanish is spoken in home	0.00	0.00	1.00	0.08 [d]	0.56 [d]
Income/SES					
Annual income in 2011 (thousands)	48.9 [ab]	34.8 [ac]	23.8 [bc]	43.7 [d]	30.5 [d]
Income to poverty ratio	2.32 [ab]	1.61 [ac]	1.03 [bc]	2.08 [d]	1.35 [d]
Number of earners in HH	1.58 [ab]	1.43 [ac]	1.32 [bc]	1.53 [d]	1.39 [d]
Family owns their own home	0.38 [b]	0.31	0.27 [b]	0.35	0.30
Family owns a car	0.86 [b]	0.80	0.75 [b]	0.83	0.78
Family does not always have enough food	0.12 [b]	0.17	0.23 [b]	0.12 [d]	0.22 [d]
Family receives public assistance	0.59 [ab]	0.70 [ac]	0.86 [bc]	0.61 [d]	0.80 [d]
High community poverty density	0.28 [b]	0.37	0.39 [b]	0.31	0.37
Household composition					
Total household members	4.69 [b]	5.00	5.19 [b]	4.79 [d]	5.09 [d]
Total household members over 18	2.28	2.30	2.47	2.33	2.37
Child lives w/two parents	0.62 [b]	0.57 [c]	0.40 [bc]	0.56	0.52
Mother was < 20 at child's birth	0.10 [a]	0.17 [ac]	0.09 [c]	0.14 [d]	0.10 [d]
Household includes grandparent of child	0.11	0.15 [c]	0.09 [c]	0.14 [d]	0.09 [d]
Household includes other adult relatives (e.g aunts/cousins)	0.22 [b]	0.28 [c]	0.43 [bc]	0.26 [d]	0.35 [d]
Any siblings in household	0.72 [ab]	0.80 [a]	0.82 [b]	0.71 [d]	0.83 [d]
Number of siblings in household	1.21 [ab]	1.43 [a]	1.50 [b]	1.24 [d]	1.48 [d]
N	960	786	1140	1241	1624

Note. Public assistance is defined as receipt of food stamps, welfare, free/reduced price lunch, or WIC (women, infant, children) benefits. High community poverty density is defined as a community where >20% of families have income to needs ratio below 1. Columns that share a superscript are significantly different from each other at p<.05. Superscripts *a,b,c* indicate differences by home language and *d* indicates differences by immigrant status.

Appendix 3A. Characteristics of Hispanic households by home language and immigrant status (continued)

	Home language			Immigrant parent	
	English only	English & Spanish	Spanish only	No	Yes
Parent work schedules					
All parents in HH worked at least 40 hrs/wk	0.18	0.19	0.18	0.20	0.18
All parents in HH worked at least 20 hrs/wk	0.33	0.30	0.28	0.34	0.28
No parents in HH worked any hours	0.28 [ab]	0.36 [a]	0.40 [b]	0.32	0.36
Mother's work schedule					
Mother worked any hours	0.51 [a]	0.43 [a]	0.42	0.51 [d]	0.42 [d]
Conditional on any hours worked:					
Total hours worked	39.01	39.44	37.53	38.91	38.59
Proportion during nonstandard hrs+	0.21 [ab]	0.32 [a]	0.32 [b]	0.24 [d]	0.31 [d]
Proportion during overnight hrs+	0.03 [a]	0.05 [a]	0.04	0.03	0.04
Proportion during weekend hrs	0.08 [ab]	0.14 [a]	0.15 [b]	0.10	0.14
Father's work schedule					
Conditional on father in household:					
Any hours worked	0.73	0.71	0.73	0.73	0.72
Conditional on any hours worked:					
Total hours worked	43.1 [a]	47.8 [a]	45.0	44.3	46.2
Proportion during nonstandard hrs	0.25	0.23	0.28	0.24	0.27
Proportion during overnight hrs	0.05	0.03	0.05	0.04 [d]	0.05 [d]
Proportion during weekend hrs	0.08	0.08	0.09	0.07 [d]	0.10 [d]
N	960	786	1140	1241	1624

Note. Columns that share a superscript are significantly different from each other at p<.05. Superscripts a,b,c indicate differences by home language and d indicates differences by immigrant status.