From Birth to School: Early Childhood Initiatives and Third-Grade Outcomes in North Carolina

Helen F. Ladd
Clara G. Muschkin
Kenneth A. Dodge

Abstract

This study examines the community-wide effects of two statewide early childhood policy initiatives in North Carolina. One initiative provides funding to improve the quality of child care services at the county level for all children between the ages of 0 to 5, and the other provides funding for preschool slots for disadvantaged four-year-olds. Differences across counties in the timing of the rollout and in the magnitude of the state financial investments per child provide the variation in programs needed to estimate their effects on schooling outcomes in third grade. We find robust positive effects of each program on third-grade test scores in both reading and math. These effects can best be explained by a combination of direct benefits for participants and spillover benefits for others. Our preferred models suggest that the combined average effects on test scores of investments in both programs at 2009 funding levels are equivalent to two to four months of instruction in grade 3. © 2013 by the Association for Public Policy Analysis and Management.

> "I believe North Carolina can do more for its children than we have done before. I believe we can do more for young children than any other state has done. And I am ready to lead a crusade for the future of our young children and our state." Jim Hunt inaugural address, 1993, quoted in Ponder (2011).

INTRODUCTION AND OVERVIEW

As part of its War on Poverty in the 1960s, the federal government began investing in early childhood programs, such as Head Start, with the goal of better preparing disadvantaged children for school. Since then, many states have provided financial support for early childhood programs. Rigorous evaluations of a few intensive, high-quality programs, such as the Carolina Abecedarian and the High Hope/Perry Preschool Project, demonstrate positive effects on program participants that persist into adulthood (Mervis, 2011). The path from small randomized controlled trials to public policy, however, is not straightforward (Gormley, 2011). Although many positive outcomes also emerge from evaluations of the federal Head Start and Early Head Start programs and various state programs, the results from studies of these larger and less intense programs are somewhat mixed (Barnett, 2011). Missing from most of the existing studies is attention to the broader effects of programs when implemented at scale and evaluation of effects on all groups of children in a

Journal of Policy Analysis and Management, Vol. 33, No. 1, 162–187 (2014)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
DOI:10.1002/pam.21734

community, not just the participants. This study of two statewide preschool programs in North Carolina is designed to help fill these gaps.

North Carolina, a large and diverse Southeastern state with 9.3 million people, has actively pursued early childhood policies since the early 1990s. Its nationally recognized Smart Start (SS) Initiative began as a demonstration project in 1993 with the goal of improving the delivery of services at the county level to all children between the ages of 0 and 5, and was gradually expanded to all counties. The More at Four (MAF) Program, approved by the legislature in 2001, was designed to supplement SS by delivering a high-quality preschool academic skills program for disadvantaged four-year-olds. Although the translation from developmental theory to statewide policy is sometimes loose, these two programs exemplify two major tenets of contemporary models of the development of early achievement gaps in education. The SS Initiative responds to the disadvantage that many children experience across the first five years of life in cognitive stimulation, secure adult relationships, nutrition, and access to health care and community resources by strengthening the child care system and improving the functioning of families. The MAF Program targets remediation in kindergarten readiness skills during the year just prior to kindergarten for a subgroup of disadvantaged children.

The purpose of this study is to examine the community-wide effects of these two initiatives on third-grade test scores, the first grade level at which test score information is available in North Carolina. Planned and haphazard differences across counties in the timing of the rollout and in the magnitude of the state financial investments relative to the populations of age-appropriate children provide the variation in programs needed to estimate their effects in subsequent years. Importantly, this variation allows us to include in our models county fixed effects, year fixed effects, and county time trends or other time-varying covariates at the county level. Thus, in our models we control for differences across counties and years that might otherwise confound the estimated effects of the programs.

The community-wide focus differentiates this study in two ways from many other studies of early childhood programs, which typically focus on how such programs affect the program participants alone. First, it avoids the statistical problem of selection bias that arises whenever participating children are not randomly assigned to the program under investigation. The bias arises because the parents who enroll their children in such programs may differ in unobservable, but relevant, ways from nonparticipating children who are in the comparison or control group. By focusing on the average effects on all children in a community, our community-wide focus avoids the selection bias related to individual parental decisions of that type because parents are highly unlikely to select into a different county before a child's birth merely to take advantage of that county's future program funding. We minimize the likelihood of other types of selection biases by including in our models county fixed effects and numerous individual-level and county-level covariates.

Second, the approach means that we are estimating the full effects of the programs—both those on the participants themselves and those that spill over to other children, in either positive or negative ways. Positive spillovers would arise, for example, if the state funding brings with it minimum quality standards that raise the quality of day care or preschool services for other children in the community not receiving direct funding under the program. In addition, once children are in elementary school, all children might benefit from being in classrooms with higher proportions of children who come to school ready to learn and with less need for remediation. Negative spillovers might arise if quality mandates reduce the number of day care or preschool slots available to children not funded by the program or, more generally, if a fixed number of openings means that some children benefit at the expense of other children. Our broader approach to the evaluation of early childhood programs generates information that should be useful for policymakers, particularly in the current fiscal environment in which states face tough budgetary trade-offs that often involve cuts in funding for preschool programs.[1]

The outcomes of interest for this paper are children's third-grade standardized test scores. Performance by the end of third grade is generally regarded as a stable indicator of student achievement and a reliable predictor of longer-term educational outcomes.

Overview of the Approach and Preview of Findings

As described in more detail below, the unit of observation is the individual child clustered by county, and the data set includes all children born in North Carolina between 1988 and 2000 who can be matched to their third-grade test scores in subsequent years. Our matched sample includes about 891,000 children from 13 birth cohorts. We estimate models of third-grade outcomes as a function of the levels of state financial allocations to the county for SS and MAF in the year(s) when the child was at the appropriate age to benefit from the presence of the state funding. Thus, the key independent variables are the dollars (per age-appropriate child) allocated to a county for each of the two programs. We do not explicitly measure differences across counties in program quality, independent of funding levels. Average shortfalls in the quality of implementation and fidelity to program goals are captured implicitly in our analysis of child outcomes. Information from North Carolina birth records permits us to control statistically for individual family and child characteristics at birth, including, for example, the education level of the child's mother and the birth weight of the child.

The inclusion of county fixed effects in all our models means that we are identifying the effects of each program based on the within-county variation across cohorts of students in the state allocations to the programs. We also include time fixed effects to control for any secular trend in student outcomes or state policies. Implicit in this identification strategy is the assumption that the introduction or funding level of the programs in a particular county year is uncorrelated with other county-year determinants of subsequent student outcomes. The possibility of such confounding factors is the biggest potential threat to the validity of our conclusions and one that needs to be carefully addressed. In our basic models, we simply do so by including county-specific time trends. In our enriched models, we substitute a large number of county- or school-specific covariates for the time trends. In all models, we take account of the potential for serial correlation in within-county test scores over time by clustering errors at the county level.

A weakness of the study is that we do not address endogenous variation in program fidelity across county-years that might affect outcomes. Nor do we address mediating factors that could inform an understanding of the processes through which a program exerts change. In theory, most state policies in early education are based on an implicit four-stage model, in which (a) the state allocates dollars to a local entity to be spent in a particular way, which (b) instigates local action such as implementing a program or recruiting children into a program; (c) changes children, presumably in the form of increased competence; finally, (d) leads to later child outcomes such as improved test scores. The current study evaluates the impact of (a) on (d) without regard to (b) or (c).

[1] Barnett et al. (2010) shows recent cutbacks in spending in many states.

The limit of this study is also its strength: It rigorously evaluates program impact at the level at which the state policymaker intends effects to occur. Although our outcome variables are measured and analyzed at the level of the individual student, the estimates of interest are the averages across all counties of the average effects within each county, across all students subject to the treatment compared to non-treated (or less well-funded) cohorts of students over the full period. The estimates are population-impact coefficients that are relevant at the policy level for guiding funding decisions.

We find substantial positive effects of each program on third-grade test scores in both math and reading. Our preferred results come from enriched models that suggest that the combined average effects on test scores of investments in both programs at 2009 funding levels are equivalent to about four months of reading and about two months of math instruction in grade 3. Moreover, consistent with policy goals, and as would be predicted by past research, the effects are somewhat larger for children whose mothers have less than a high school education compared to those whose mothers have more education. Thus, the programs narrow the parental-education achievement gap for third graders. Given the mother's education, the effects do not differ between children of black and white mothers, and the effects are somewhat mixed for children of Hispanic mothers relative to those with white mothers. Combined with other effects of the two programs, some of which are documented in separate research, the effects on third-grade test scores reported in this paper indicate that the two programs appear to be cost-beneficial investments for the state.

Previous Research

The High Scope/Perry Preschool and the Carolina Abecedarian Project, which were initiated over 40 years ago, have received great attention in the early childhood literature (see the special section of *Science* devoted to early childhood programs, August 19, 2011), reflecting their high-quality and careful designs. Even more importantly, the programs were designed as experiments in which poor families were first recruited into a study and then were randomly assigned to either the treatment group of children who participated in the program or the control group. Although the random assignment generates strong internal validity, the results can at best be generalized only to the types of families that were sought and successfully recruited into the programs. In the case of the Perry Preschool Program, exclusion criteria ruled out children of working mothers because their home visits, which were an essential part of the program design, could not accommodate mothers who worked outside the home. Thus, the Perry Program findings might not generalize to contemporary society, in which virtually all able-bodied mothers of preschoolers try to work outside the home at least part-time.

To avoid the external validity problems of random-assignment experiments, researchers have creatively turned to a variety of observational data sets and nonexperimental techniques. These techniques include comparing outcomes for program participants in Chicago's Child Parent Centers to those for demographically similar children in the same neighborhoods (Reynolds et al., 2002); evaluating the federal Head Start Program by using national data and controlling for family fixed effects so that participating siblings are compared to nonparticipating siblings (Currie & Thomas, 1995); evaluating the effect of universal prekindergarten in Georgia on school outcomes using a difference-in-differences analysis of NAEP data (Fitzpatrick, 2008); and using a regression discontinuity design to analyze the long-term effects of the 1965 Head Start rollout for residents in counties on either side of the cutoff point for assistance (Ludwig & Miller, 2007). Most nonexperimental studies demonstrate that the programs have positive effects, although not always for all outcomes or for all groups, and not always for extended periods.[2] In the Currie and Thomas (1995) study of Head Start, for example, the authors find short-run gains in test scores for both white and black children, but the gains for black children fade quite quickly. In addition, the authors find that Head Start reduced grade repetition for white, but not for black, children.

Two recent regression discontinuity studies focus specifically on state preschool programs. In the first study, Gormley et al. (2005) analyzed Oklahoma's universal voluntary preschool program for all four-year-olds in the state. Making use of the state's strict age cutoff for program eligibility, the authors compared the cognitive skills of the children who just made the cutoff one year, and hence who participated in the program, with those of approximately similar age who were not eligible to participate. The authors found that the program generated relatively large gains for various domains of the Woodcock Johnson Achievement test and that all racial and income groups appear to have benefitted from the program. Using a similar methodology to evaluate preschool programs in five states, Wong et al. (2008) also find positive effects on various cognitive measures, but the magnitudes differ across the states. The study design is limited in that the authors can focus only on very short-term outcomes, namely those that can be measured at the end of the program, because the control group gains access to the program the following year. Thus, such studies can shed no light on whether the program benefits persist or fade over time or on the nature of benefits that might arise only several years after participation in the program.

The present study contributes to the literature by addressing the relation between county funding levels and cognitive outcomes for children over the medium term (minimum of four years). It differs from the work just described not only by its focus on longer term outcomes, but also by its attention to how early childhood initiatives in NC affect outcomes for all children in the affected counties, not just the participants in specific components of the initiatives. In that sense, this study is comparable to three other recent studies of government investments in early childhood programs.

The first study examines the effects of child care subsidies in Norway (Havenes & Mogstad, 2011). Using individual-level data from the Norwegian census, the authors examine the effects of large-scale increases in government investments in some municipalities on subsequent outcomes for children who were living in those municipalities at the time. The study differs from the present one in that government investments were more diffuse than in our case and in assessing long-term outcomes extending far beyond the school years. For the children who had the potential to benefit, the authors find that the subsidies increased their educational attainment and labor market participation, and reduced their dependence on welfare. Additional analyses in that study show that girls and children with low-educated mothers benefited the most from the child care subsidies.

Far less positive impacts for children emerge from a study of child care subsidies in Quebec (Baker, Gruber, & Milligan, 2008). Although the main purpose of the study was to examine how the subsidies affected maternal labor supply, the researchers' use of longitudinal survey data allowed them to examine how child care subsidies and publicly financed preschools affected child and parental well-being and parental practices, as reported by parents. The authors find that the subsidized child care,

[2] For a recent meta-analysis of 84 studies and careful discussion of the relevant developmental literature that bears on the effectiveness of early childhood and preschool programs, see Duncan and Magnuson (2013).

which induced a striking increase in nonparental child care and greater employment of women, led to a number of negative outcomes for children such as anxiety, aggressiveness, slower development of motor and social skills, poor child health status, and illness. The rigor of the study notwithstanding, the authors are careful to highlight two limitations. First, their results apply only to married (or cohabiting) mothers and, second, the measured adverse effects on children are short-term. The authors note that they "cannot rule out the possibility that our findings represent a short-term adjustment to child care, and not a long-run negative impact" (Baker, Gruber, & Milligan, 2008, p. 713).

Finally, a careful historical study of state grants for kindergarten programs at the district level during the 1960s and 1970s generates mixed results at best (Cascio, 2009). Using individual-level data from the U.S. Census Public Microdata Samples (PUMS), Cascio finds that, on average, white children who turned 5 after the state reform exhibited lower rates of high school dropout and of institutionalization as adults, but finds no other positive effects for a range of possible outcomes. Despite the finding that kindergarten enrollments of black children increased at similar rates to enrollments of white children in response to the state grants, no other positive effects emerged for blacks. The author argues that the state grants for kindergarten programs may have had the effect of crowding out participation in higher quality federally funded early education programs among many low-income black children. The possibility of crowd out suggests that impacts of a program might differ between those observed during a small-scale trial and those observed when disseminated at scale in a broader community context.

SS AND MAF INITIATIVES

Both North Carolina initiatives take the form of state funding for local communities. Especially in the case of SS, local communities have discretion in how they spend the money. For both initiatives, we view the treatment of interest to be the availability of funds at the county level during the years that a child was age-eligible for the programs.

SS Initiative

The SS Initiative was the centerpiece of Governor Jim Hunt's "crusade" for the future of North Carolina children, referenced in the quotation that begins this paper. Indeed, with the governor's strong and energetic leadership, the state accomplished what no other state had yet done. Specifically, in 1993, North Carolina established a comprehensive, collaborative, and decentralized initiative designed to assure that all the state's children and their families have access to high-quality services that would prepare them for school. Services were to be available for all children aged 0 to 5, not just to those with low family income.

The initiative combined a top-down comprehensive systems approach designed to address a broad set of challenges facing children in child care and education, health, and family support, with a bottom-up implementation approach designed to maximize local ownership and commitment. The approach was collaborative in that it required representatives from a wide range of groups, including families, to be involved at all levels of the initiative, with the focus throughout on partnership and teamwork. It was decentralized in that the state provided funding to broadly constituted county-level (and in some cases multicounty) partnerships setup specifically to implement the initiative. Not until 1996 did the state impose any specific restrictions on how the funds were to be used. At that time, the legislature required that at least 30 percent of the funding be devoted to child care subsidies. In the following year it mandated that in the aggregate, at least 70 percent was to be spent on child care related activities.[3] Subsequently, the state set up a performance monitoring system with five performance goals in the areas of access to quality child care, health care, and support for families to raise healthy children who are prepared for school (Ponder, 2011). Because of the intentional strategy to encourage local decisionmaking, the implemented program can best be described as a pool of financial resources with guidelines for spending. By design, that pool varied over time and across counties as the program grew and state resources evolved. Although this somewhat loosely defined policy limits scholars' ability to identify program mechanisms precisely, it is a realistic model for how state-level funding often operates.

In the spirit of public–private partnership, the legislature established a new non-profit organization called the North Carolina Partnership for Children, Inc. (NCPC) to support and oversee the local partnerships. The legislature also called for 10 years of evaluation that was contracted out to the University of North Carolina's Frank Porter Graham Child Development Institute (FPG).

The characteristics of top-down guidelines and decentralized implementation make the SS Initiative both difficult to define precisely and even more difficult to evaluate. In an early evaluation, for example, FPG found that the 12 initial partnerships had funded 244 individual local service activities (Bryant et al., 2002, 2003). The evaluation team grouped activities into three program categories: child care quality, family functioning, and children's health—along with a process category that included activities designed to improve interagency collaboration. They then developed a logic model to link each category to expected short-term and long-term goals. Funding for the two main categories—raising the quality of child care and improving the functioning of families—was intended to promote the central goal of school readiness by developing skills and improving behavior; funding for health care promoted the short-term goal of increasing immunizations and expanding access to regular checkups, toward the longer-term goal of better child health; and attention to interagency processes was intended to promote child well-being by providing more coordination, eliminating inefficiencies, and reducing service gaps for children and their families.

Over a 10-year period, evaluators at FPG produced 35 separate studies. Many provided valuable information on how the initiative was being implemented, and others provided evidence related to short-term outcomes.[4] FPG studies focused on two main impacts: (1) the impact of SS on the quality of child care services, and (2) the impact of SS-funded center quality on the school readiness of participating children. These impacts were particularly germane to the initiative's goal of ensuring "that all children enter school healthy and prepared to succeed," primarily through improving the quality of early care and education for children ages 0 to 5 (Bryant et al., 2003).

To examine the impact on the quality of child care services in North Carolina over time, Bryant et al. (2002) collected information from centers receiving funds from the initial 12 SS partnerships in 18 counties, over three waves of onsite observations between 1994 and 1999. They concluded that numerous child care improvements were related to center participation in SS, including better classroom quality, increases in teacher qualifications, and increases in the percentage of centers licensed at higher levels and with national accreditation. Although the authors

[3] This mandate was not constraining in that more than 75 percent was already being spent on those activities. In another adjustment, the legislature required counties to use up to 50 percent of their Smart Start funds for child care subsidies if the county had a waiting list that could not be met in other ways (Ponder, 2011).

[4] http://www.fpg.unc.edu/projects/evaluation-north-carolinas-smart-start-initiative.

cannot conclude with certainty that SS was the cause of these quality improvements, changes of this type are likely one of the mechanisms through which the initiative may affect long-term academic achievement.

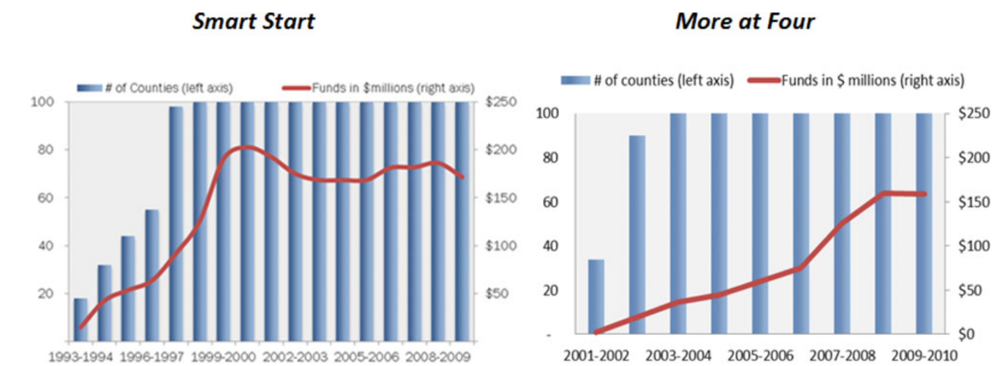
Similar issues emerge in FPG studies that examine the benefits for children who participate directly in centers receiving SS funding. Maxwell, Bryant, and Miller-Johnson (1999) compared the language and social skills of children who had attended child care centers participating in six SS partnerships with those of a comparison sample within the same kindergarten classrooms. They found higher cognitive (but not social) skills for children who attended SS-funded child care centers that had received quality improvement, such as onsite technical assistance and higher levels of teacher education. One limit of this study design is that it assumes no spillover effects. That is, it assumes that control children (taken from the same kindergarten classrooms as SS children) were not affected by program funding. If there is positive spillover, then the FPG analyses will have underestimated the positive impact of SS.

A subsequent study combines research questions, evaluating both the change in child care quality across centers in 20 partnerships and the extent to which attending a higher quality program predicts school readiness for individual children (Bryant et al., 2003). This study included child care quality measures, teacher ratings, and direct assessment of cognitive and social skills. The researchers found that the quality of the individual centers in the sample had increased over time, and that children attending higher quality classrooms had higher levels of kindergarten readiness. Although this finding held across the sample, causal claims were not warranted because there was no comparison or control group of children enrolled in centers not directly benefiting from SS funding.

Taken as a group, the FPG studies suggest that SS was associated with positive effects on center quality and child school readiness, but they do not provide a convincing causal link either at the center or the individual student levels. Moreover, the breadth of the SS Initiative, combined with variations in the way it was implemented at the local level, made it difficult for the FPG researchers to provide an overall evaluation. Nonetheless, the FPG studies were sufficiently positive to generate ongoing legislative support and for the initiative to win a number of national awards, including the prestigious Ford Foundation-Harvard award for Innovations in American Government, to attract attention from other states, and to contribute to a national movement oriented toward a comprehensive approach to the needs of young children. The studies also provide a basis for our hypothesis that the SS Initiative generated medium-term effects on children's academic outcomes, which are the focus of this study.

The initiative was started in 1993 in 12 pilot partnerships that represent 18 of the 100 North Carolina counties. The pilot counties were selected by independent experts and were specifically chosen to be representative of North Carolina's diversity and geography, with one recipient from each congressional district.[5] As shown by the left-hand graph in Figure 1, the program increased to more than 50 counties by 1997 and to all 100 by the 1998 to 1999 school year (left axis). During that period, SS funding rose to a peak in 2000 of $250 million (in 2009 dollars) and has since fallen.

[5] The 12 experts used county-level data on the number of families in poverty, tax base, number of children under age 5 with working parents, and number of children with special needs to rank all 100 NC counties into one of four resource bands: high, medium high, medium low, and low. For each resource band, three counties (or in one case a region of seven small counties) were chosen as required by the inaugural legislation.



Notes:
1. *Data Sources*
   (a) Yearly Smart Start Funding data provided by North Carolina Partnership for Children FY 1993 to 2009, NC Division of Child Development FY 1993 to 2009.
   (b) Yearly More at Four Funding data provided by North Carolina Office of Early Learning.
   (c) Monthly CPI data provided by Bureau of Labor Statistics, U.S. Department of Labor.
2. Both figures are in June 2009 dollars using the CPI as of July in each year as an inflator.

Figure 1. North Carolina Early Childhood Initiatives.

Funding per 0- to 5-year-old child varied across counties and over time. Per child funding peaked in 2000 at about $400 per 0- to 5-year-old child and has since fallen to about $220 per child in 2009. A child living in a county with SS funding for all five of his or her early childhood years would have access to SS funding equal to about five times these per child amounts. (See online Appendix Figs. A1 and A2 for the variation in funding across counties by year and for maps illustrating the geographic variation in program funding for selected years.[6]) We use this variation in the timing of the introduction of the program and in its intensity across counties to identify the effects of the initiative.

More at Four

When Governor Mike Easley took office in 2001, the state added the MAF Pre-Kindergarten Initiative on top of SS. The logic was that even with community-wide efforts to improve general child care and child health, many low-income children would still be deficient in cognitive skills in the year prior to kindergarten matriculation. A key goal of the program was to make high-quality preschool settings available to all children by ensuring access to disadvantaged children.

Like SS, this initiative provided state funding for a top-down program that was to be administered at the local level. Specifically, MAF provided funding for slots for eligible four-year-old children in pre-k centers. Compared to SS, the state was far more prescriptive with respect to how the funds were to be used. First, they were to be used only for "at-risk" four-year-olds, with priority given to children not currently served in a formal preschool or child care program. Eligibility was based on the poverty status of the child and specific risk factors defined in the legislation.[7] Second, the funds were to pay for slots in settings operated only by

[6] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

[7] Depending on family size, all children meeting the age requirement in families with annual incomes at or below 75 percent of the state median income are eligible. In addition, an age-appropriate child

high-quality providers. While local administrators were free to allocate the funded slots to a variety of providers including public schools, private for-profit and non-profit child care centers, or Head Start programs, any recipient agency of MAF funds was required either to meet state-delineated quality standards or to be on track to meet them.[8] These standards refer to staff qualifications, class size, teacher–child ratios, and North Carolina child care licensing requirements. Through this requirement, the state's goal was to promote high-quality preschool not only for the funded children, but also for other children enrolled in the same centers. About one-third of children in MAF classrooms were not directly funded by MAF (Peisner-Feinberg, Elander, & Maris, 2006); hence, roughly 50 percent more children than were directly funded potentially received spillover benefits in the form of higher quality classrooms in MAF-funded preschool settings. Spillover benefits for peers in the kindergarten and later classrooms of MAF recipients are plausible as well.

In recognition of its educational goals, namely to provide a high-quality classroom experience for the state's at-risk four-year-olds, the program was overseen and managed for 10 years by the Office of Early Learning within the North Carolina Department of Public Instruction. To avoid competition or overlap with SS, local committees charged with making decisions about the allocation of MAF slots were jointly convened by the chair of the local SS partnership and the superintendent of schools in each county. Once again the Frank Porter Graham center was the official evaluator. MAF began with funding for 32 counties in 2002 and was expanded to all counties within a few years.[9]

Early FPG studies provided descriptive information on how the program operated. The median class size for MAF-funded students in 2004 to 2005 was 18, with about two-thirds of the students in each classroom funded by the program. Consistent with the program guidelines, about 90 percent of the children qualified for free or reduced-price lunch and more than three-quarters were not receiving services at the time of enrollment. About a third were African American, a third were white, and about 20 percent were Latino. As measured by staff credentials, the quality of the programs was generally high, but higher in the centers operated by public schools than those run by community for-profit and nonprofit centers. For example, among the lead teachers in the public school settings, 99 percent had a bachelor's degree or higher and 75 percent had an appropriate teaching license, compared to 65 percent and 15 percent, respectively, among the lead teachers in community settings (Peisner-Feinberg, Elander, & Maris, 2006).

Classroom observations, parental surveys, and samples of students provided information on the quality of classroom practices and short-term child outcomes. FPG evaluations of child outcomes focused on measuring gains for MAF-funded children during the year that they participated in the program. The studies show positive gains in measured skills beyond what would be expected for this target population. Furthermore, children who entered the program at higher levels of risk exhibited greater gains during the preschool year on an array of cognitive and behavioral outcomes. At the same time, the absence of a comparison group made it difficult for the researchers to make strong causal claims.

whose family has higher income may be eligible if the child is limited English proficient, is disabled, has a chronic illness, or has a developmental need. Service provision to eligible children is limited by the available resources.

[8] In 2004 to 2005, 50 percent of the slots were in public school settings, 31 percent in for-profit community child care centers, 11 percent in nonprofit child care centers, and about 10 percent were in Head Start Programs (Peisner-Feinberg et al., 2006, p. 9).

[9] It continued until the legislature moved the program to the Department of Health and Human Services in 2011 and renamed it the NC Pre-K program.

In a subsequent, more ambitious study, the evaluators examined the effects of the program on the longer-term outcomes of third-grade math and reading scores (Peisner-Feinberg & Schaff, 2010). Using two early cohorts of MAF participants, the authors linked the participants to their third-grade test scores, and then compared the test scores of the participants to a demographically similar comparison group of third graders within the same counties in the same years (2007 and 2008). In these analyses, the authors controlled for county-specific levels of state and local education spending. The main finding was that the program participants who were from low-income families exhibited modestly higher third-grade test scores than their equally impoverished same-county counterparts who did not participate, with effect sizes that ranged from about 0.10 to 0.14.[10] Although the authors interpreted this finding to mean that the program was serving its primary policy goal of reducing income-related achievement gaps, the results are still at best suggestive. The researchers' use of a matched comparison group rather than a randomly assigned control group means that they were not able to control for the selection effects that bedevil this type of evaluation. As a result, their impact estimates could be inflated. At the same time, to the extent that positive spillover effects operated as program planners hoped, such that matched-comparison children received indirect benefits, the FPG evaluation would underestimate the true program impact.

In recognition of the potential spillovers, we use the same approach as for SS and define the treatment as the availability and level of MAF funding (per age-appropriate child) in the county. The right graph in Figure 1 shows the rapid rollout of the program across counties and the total funding levels over time. (For more information on the variation across counties in the level of funding and availability of classroom slots across counties, see the online Appendix Figs. A3–A5.[11]) Because SS funding was available (at varying levels) in all counties when MAF was rolled out, all our estimates of the effects at MAF are conditional on the presence of SS programs.

ANALYTICAL APPROACH

Consistent with the previous discussion of the treatment, the relevant treatment groups for SS or MAF are all the children of the appropriate age living in each county when SS or MAF funding was available in the county. Our analytic approach is to estimate relatively straightforward models that exploit the variation by county and over time in the introduction and per-child funding of the two programs.

The basic model takes the following form:

$$O_{icbt} = \beta_1 SS_{ict}{}^* + \beta_2 MF_{ict}{}^* + \beta_3 X_{ib} + \beta_4 Y_{it} + \beta_5 T_c + \alpha_c + \gamma_b + \varepsilon_{icbt},$$

where O_{icbt} is a third-grade outcome, such as a math or reading score, in year t for the ith student born in county c in year b, $SS_{icb}{}^*$ is a SS variable for the ith child, as defined below, $MF_{icb}{}^*$ is a MAF variable or the ith, as defined below, X_{ib} is a vector of characteristics of the ith child and his or her parents children at the time of birth including, for example, the child's birth weight and the education level of the child's mother, and Y_{it} are characteristics of the ith child observed in year t, such as the race of the child as reported on school records.

[10] The findings were less positive for the group of participants whose families were not poor, but had participated because they met eligibility requirements by being at-risk in another way.

[11] All appendices are available at the end of this article as it appears in JPAM on-line. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

The model includes county fixed effects (α_c) and year fixed effects (γ_b), as well as county-specific linear time trends (T_c), based on year of birth. The inclusion of fixed effects for each county rules out the possibility that the estimated effects of the early childhood programs are the result of other activities or investments that differ systematically across counties in time-invariant ways that are correlated with the early childhood investments. The time trends are intended to control for unmeasured factors such as changing demographic or economic characteristics of a county that may affect county-specific trends in test scores. As a result, our estimates of program effects reflect departures from test score trends within each county.[12] In subsequent models discussed below, we substitute a number of time-varying covariates at the county level for the county time trends. Finally, we include year fixed effects to control for state-level factors that could potentially affect all counties in a particular year, such as budget conditions or changes in the economy that may affect the extent to which families and their children are stressed.[13]

Of primary interest are the coefficients on the two program variables. In the basic model we define the SS variable as

$$SS_{icb}{}^{*} = \Sigma PSS_{icb(\text{age}=0,\ldots,4)},$$

where PSS refers to the penetration of SS defined as the inflation-adjusted dollars per child aged 0 to 4 in the relevant county in each of the years when the child was under 5. The child is eligible prior to entering school, and we can precisely determine eligibility using the child's birth date. As the sum of the penetration rates over the years when the child was under 5, this measure accounts not only for different spending per year, but also for the possibility that funding may not have been available in the relevant county during all the child's preschool years. The major determinants of within-country changes over time in this variable are changes in budgetary and allocation decisions at the state level.

The analogous measure for $MF_{icb}{}^{*}$ is $PMF_{icb(\text{age}\,=\,4)}$. In this case, the penetration rate refers to the inflation-adjusted dollars per four-year old for MAF in the relevant county for a single year, namely the year in which the child was four years old. Once again we use the child's birth date to determine eligibility, paying close attention to the relationship between the month of birth and the month at which eligibility starts or ends. Note that the variable is normalized by the total number of four-year-olds in a county, not by the smaller number of four-year-olds who meet the eligibility criterion. To the extent that funding allocations are determined, at least in part, by the number of eligible four-year-olds in a county, this variable would vary across counties in line with the proportion of at-risk four-year-olds to all four-year-olds. Much of the variation of this type would be absorbed, however, by the county fixed effects. The county time trends control—albeit imperfectly—for within-county trends over time in the proportion of eligible children to all four-year-olds.

In the results sections below, we describe several variations of this basic model designed to explore its strengths and weaknesses. We then enrich the model by substituting a number of time-varying variables at the county and school level for

[12] Besley and Burgess (2004) show that allowing for differential trends between areas in a difference-in-difference regression may eliminate otherwise significant and large treatment effects. That does not seem to be the case for our model. Wolfers (2006), however, argues that in some situations, especially those in which there is a dynamic interaction between stocks and flows, area-specific trends may generate misleading results. As we note below, there are also other reasons to pay close attention to the time-varying patterns at the county level.
[13] Ananat, Gasman-Pines, and Gibson-Davis (2011) find significant community-level influences of large-scale job losses through plant closings in North Carolina on test scores among fourth and eighth graders.

the county time trends. In addition, we test for differential effects by population subgroups.

DATA

We base our analysis on four data sets. The first is from the North Carolina Division of Public Health (Vital Records) and provides individual records on all births in North Carolina from 1988 to 2000. Reflecting the state's growth, the number of births increased from 95,000 in 1988 to 119,000 in 2000. These birth records include the infant's date of birth, gender, birth weight, marital status of the mother, education level of the mother, whether a father can be identified, race, and county of origin of the mother.[14]

The second data set includes rich administrative information from the NC Department of Public Instruction made available to us through the North Carolina Education Research Data Center, housed at Duke University. The key variables in this data set are third-grade reading and math standardized test scores. These scores are based on state-mandated tests linked to the statewide curriculum and are used in the accountability systems to which all North Carolina schools are subject. Because of changes in the scales of the tests over time, we normalize all test score results by subject, year, and grade. Hence, the unit for all the test scores is a standard deviation, and all effects are measured relative to the overall distribution of test scores. From this data set, we also use race of the child as recorded in grade 3, which sometimes differs from the race of the mother in the birth records.[15]

The third data set includes annual funding levels by county for each of the two programs provided in the form of administrative records by the programs. We have inflated all funding levels to 2009 dollars. The fourth includes publicly available demographic and school quality information at the county and school levels over time.

Crucial to our analysis is our ability to match birth records to children's third-grade records in the North Carolina education data. Table 1 shows that the total match rate for all births from 1988 to 2000 is about 68 percent. The missing 32 percent could reflect movement of families out of the state, children opting out of the public school system in favor of private schools, or errors in our matching process. Using data from the American Community Survey, we estimate that of all children born in North Carolina, about 74 percent attend public schools in the state at the age of 8 or 9. Hence, our match rate of 68 percent accounts for more than 90 percent of the relevant births.

Potentially more problematic is the variation in match rates across racial groups. The matching rate for children of white mothers who account for 63 percent of the births over the period is about 66 percent, close to the overall average. At 76 percent and 79 percent, respectively, the matching rates for children of black and Native American mothers are well above the average, which seems plausible in light of the lesser tendency of such groups to leave the state or to attend private schools. Of potential concern is the relatively low 56 percent match rate for children of Hispanic mothers. We note, though, that that Hispanics account for only 8 percent of all births, albeit a percentage that has been rising in the last two decades. On the

[14] It also includes other potentially useful information that we have not used in this study, such as gestational age, Apgar scores, and maternal behaviors such as smoking and drinking.
[15] The birth of the mother and the race of the child are highly but not perfectly correlated. For children of white mothers, for example, 95 percent of the children are identified as white in school records. For some of the sensitivity tests reported below, we also use data from this source on the eligibility of a child for a subsidized lunch in grade 3, and the racial mix of the student's peers.

Table 1. Match rates by race of mother (proportion).

	Share of births	Match to any NC county	Match to same county as birth
Total births	1.00	0.68	0.54
White non-Hispanic	0.63	0.66	0.51
Black	0.26	0.76	0.62
Hispanic	0.08	0.56	0.43
Asian	0.02	0.51	0.40
Native American	0.01	0.79	0.66
Other	0.00	0.47	0.39

Source: Match rates between birth records and third-grade test scores by race of mother for birth years, 1988 to 2000, calculated by the authors.

one hand, the low match rate may correctly reflect the far greater mobility of this subgroup of the population, given that North Carolina is a new destination state for Hispanics and many who first arrive here may not remain. On the other hand, the low match rate could reflect errors in our matching process, given inconsistency in spelling of Hispanic names across data sets. In one of our models below, we test the sensitivity or our results to the match rate for Hispanic births.

Even with the more than 90 percent overall match rate, our estimated program effects could be biased if the probability of being missing from the sample were systematically related to the funding levels of either of the programs. We test for such a relation based on the full set of births using a linear probability model similar to the basic model but in which the dependent variable takes on the value of 1 if the birth is matched to third-grade test scores and 0 if it is not. The finding that the estimated coefficients of both program variables are very small and far from statistically significant alleviates concerns about such bias.[16]

A final sample-related consideration is that we cannot be sure that all children born in a county and matched to a third-grade test were living in the birth county during their early childhood years, yet that is the assumption we make throughout the analysis. As shown in the final column of Table 1, only about 54 percent of the births can be matched to their third-grade test scores in the same county in which they were born. For this group, we can be reasonably confident that the children lived in their birth county during their preschool years. For the others, whether they lived in their birth county or the county of their third-grade schooling (or possibly some other county) depends on when their family moved from one county to another. Nonetheless, we assume that they were all subject to the relevant treatment in their county of birth. It is important to include such children in the analysis so as not to bias the result by the differential attrition from the sample of families whose moving decision might have been influenced by the existence or generosity of either of the programs.

Descriptive statistics for all the variables in the model are reported in Table 2. The individual-level control variables include characteristics of the infant including birth weight, gender, and an indicator for first born; characteristics of the mother at the time of the infant's birth, including years of education, marital and immigrant

[16] The coefficient of the More at Four variables is 0.0059 with a standard error of 0.0049, and the coefficient of the Smart Start variables is −0.008 with a standard effort of 0.0014. The sample size of the regression is 1,323,052.

Table 2. Variables in the basic and enriched models (894,979 observations).

	Mean	Standard deviation
Average third-grade test scores		
Reading	−0.017	0.995
Math	−0.017	0.994
Program variables (nonzero values only)		
Smart Start ($100 per 0- to 5-year old)	11.5	8.56
More at Four	3.34	2.52
Individual-level control variables		
Extremely low birth weight (less than 2.2 lbs)	0.005	0.070
Very low birth weight (less than 3.3 lbs more than 2.2 lbs)	0.008	0.088
Low birth weight (less than 5.5 lbs more than 3.3 lbs)	0.066	0.240
High birth weight (more than 5.5 lbs)	0.103	0.340
Female infant	0.488	0.500
Mother's education (years)	12.71	2.500
Marital status (married = 1)	0.70	0.460
Mother's age	26.04	5.870
No dad information	0.133	0.339
Mother immigrant	0.075	0.264
First born	0.443	0.497
Mother black	0.272	0.445
Mother Native American	0.014	0.119
Mother Asian	0.015	0.122
Mother Hispanic	0.046	0.210
Mother other	0.001	0.028
Child black	0.205	0.404
Child Native American	0.012	0.111
Child Asian	0.007	0.082
Child Hispanic	0.026	0.159
County demographic data, birth year		
Births to black mothers (share of births)	0.272	0.143
Births to Hispanic mothers (share of births)	0.046	0.046
Births to low-education mothers (share of births)	0.216	0.067
Population of food stamps (share of population)	0.074	0.037
Number of births (log)	7.54	1.06
Total population (log)	11.78	0.99
Median family income (2009 dollars)	57,658	10,660
Per-pupil spending by source, test year		
Federal (2009 dollars)	774	300
State (2009 dollars)	5,287	542
Local (2009 dollars)	1,883	618
School characteristics, test year		
Black students (share of students)	0.302	0.243
Other minority students (share of students)	0.096	0.123
Inexperienced teachers (share of teachers)	0.111	0.110
Average licensure score of teachers (standardized)	−0.030	0.315

status, racial or ethnic group; and data from the school records on the race or ethnicity of the child. No income measure is included among the list of control variables. One reason is that parental income is not available in the birth records. Another is that the measure typically used as a crude proxy for income in education studies, namely, eligibility for free or reduced-price lunch, is missing for some years for some cohorts of third graders in our sample and would require that we impute some of the data. A final reason is that, as a third-grade variable, it is somewhat less exogenous than the other control variables that, other than the race or ethnicity

Table 3. Variations on the basic model, selected coefficients.

	Basic model	Without individual controls	With no trends	Truncated sample	With all Hispanic births
Panel A: Reading					
Smart Start ($00s)	0.0094***	0.0084***	0.0065***	0.0167***	0.0091***
	(0.0012)	(0.0011)	(0.0011)	(0.0028)	(0.0011)
More at Four ($00s)	0.0187***	0.0166***	0.0083***	0.0083***	0.0175***
	(0.0032)	(0.0034)	(0.0024)	(0.0033)	(0.0032)
Observations	886,783	886,783	886,783	[a]	915,396
Panel B: Math					
Smart Start ($00s)	0.0081***	0.0069***	0.0052***	0.0113***	0.0078***
	(0.0016)	(0.0016)	(0.0014)	−0.0039	(0.0015)
More at Four ($00s)	0.0214***	0.0192***	0.0059**	0.0107***	0.0203***
	(0.0040)	(0.0041)	(0.0027)	(0.0032)	(0.0039)
Observations	890,827	890,827	890,827	[a]	919,212

Notes: Selected coefficients are from full regression models that include controls for a large number of child and parent characteristics at time of birth, child racial characteristics in grade 3, fixed effects for county and year of birth, and county time trends. The full equations for the basic model are included in the online Appendix, Table A1.[18] Standard errors clustered at the county level are in parentheses.

[a]The sample sizes for the truncated models differ by program and by subject. For reading, the sample sizes are 116,323 for Smart Start and 134,449 for More at Four, and for math they are 116,779 and 134,998.

***Statistical significance at the 0.01 level.

**Statistical significance at the 0.05 level.

of the child, are all measured at time of birth. Consequently, our only measure of the socioeconomic status of the child's family is the education level of the mother at the time of the child's birth. Fortunately, analyses of the Panel Study of Income Dynamics indicate that parents' education is more highly predictive of test scores than is family income and that any effects of family income are mediated through parents' expectations for the child (Davis-Kean, 2005).[17]

The county- and school-level variables in the bottom three panels of Table 2 are discussed below in the context of the enriched models.

RESULTS FROM BASIC AND ENRICHED MODELS

The main programmatic results for the basic model for reading and math test scores are summarized in column 1 of Table 3, with the full equations provided in the online Appendix, Table A1.[18] The sample sizes are slightly larger for the math models than for the reading models because more third graders took the math than the reading tests. As shown in the table in online Appendix, all the individual control variables consistently enter both models with the predicted signs. Low birth weight, having a mother who is an immigrant or is black, Hispanic, or Native American, for example, are all predictive of lower test scores. Moreover, the comparable racial or ethnic

[17] Moreover, the results are not sensitive to the inclusion of a measure of eligibility for free and reduced-price lunch. See footnote 20.

[18] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

measures from the third-grade child data also predict lower test scores. Likewise, being the first-born child, or having a mother with higher education or who is married or older, is associated with higher test scores. In most of the regressions, we specify mother's education as years of education. In the interaction models reported in Table 5, however, we convert the measure to a binary variable, with 1 indicating that the mother has less than a high school education.

Turning now to the coefficients for the program funding variables in the first column of Table 3, we find the coefficients are positive and statistically significant for each program in both subjects. For reasons we explore empirically below, however, we believe these estimates of program effects are too large. Nonetheless, we provide the following interpretation of the coefficients to help the reader put the magnitudes of these and subsequent smaller estimates into perspective.

The units of both program variables are hundreds of dollars per child. To interpret the magnitudes of the SS coefficients, consider a five-year funding level of $1,100 (or about $220 per 0- to 5-year-old child per year). This level is close to the average funding for the sample period (calculated using the nonzero funding amounts in the basic sample as shown in Table 2) and also is approximately the average funding level in 2009. Based on the reported coefficients for SS, an investment of this magnitude would generate an average gain in subsequent third-grade reading scores of 0.10 standard deviations (i.e., the reported coefficient of SS in Table 3, column 1 of panel A times 11) and in third-grade math scores (see Table 3, panel B) of 0.09 standard deviations, or about a 10th of a standard deviation in each case.

In the case of MAF, a state investment of about $300 per four-year old (the average funding level during the relevant time period for our sample) would translate into an average gain of 0.056 standard deviations (0.0187 times 3) in subsequent third-grade reading scores and an average gain of 0.064 in subsequent third-grade math scores. This $300 average, however, is far below the 2009 funding level for MAF, which is closer to $1,100 per four-year old. Based on that higher funding level, the estimated coefficients would translate into a 0.206 standard deviation increase in reading and a 0.235 increase in math.

To put these magnitudes in perspective, the typical annual gain in test scores in third grade is about 0.5 standard deviations in reading and between 0.6 and 0.9 standard deviations in math depending on the year.[19] Hence, if we assume a 10-month school year, an effect size of 0.1 standard deviation represents about two months of instruction in reading and about 1.5 months in math. Thus, according to these estimates, the two programs together at 2009 funding levels (about $1,100 per child for each) would generate gains of about four to six months of instruction. These estimated effects seem surprisingly large and hence are worth investigating in more detail.

In a true experiment, the program variables would be uncorrelated with any pretreatment characteristics. Stated differently, if the variations in program funding over time within countries were uncorrelated with the characteristics of individual children, the inclusion of the individual control variables should have little or no influence on the estimated coefficients of the program variables. In column 2, we test this proposition by reporting results for models that exclude the individual demographic control variables. Although the estimated coefficients are all smaller than those in column 1, they are the same order of magnitude and continue to be statistically significant, indicating that correlation of the program variables with the individual-level control variables is not a major concern.

[19] These estimates are based on our own calculations as well as information reported in the North Carolina State Testing Results, 2008 to 2009.

Of greater potential concern is the linear specification of the county-specific time trends. In column 3, we explore the role of these trends by excluding them from the basic model (while including the individual demographic controls). In this new specification, all the estimated program coefficients are substantially smaller than they were in the basic model, although they are still statistically significant, with the reductions being largest for the MAF coefficients. These results imply that assumptions about the nature of the population and growth trajectories at the county level are consequential for the estimation of program effects and indicate that it would be worthwhile to pay more attention to the nature of those trends. Although it would be possible to test different functional forms of the county time trends, any efforts along those lines are not likely to address the primary concern, which is that the basic model may not control adequately for other determinants of test scores that vary over time within counties and that may be correlated with the program funding variables. Hence, our preferred strategy is to replace the county time trends with a variety of such confounders.

Before doing so, we briefly refer the reader to the final two columns of Table 3 with results based on different samples of children. The results in column 4 limit the sample to the cohorts that were the appropriate age just before and just after the county in which they were born received initial funding for the program. Thus, for SS, the sample includes only those children in a county who were six years old when the program was introduced, and hence were too old to benefit, and those who were born the following year, and hence could benefit fully. Analogously, for MAF, the sample includes children who were 5 and those who were 3 when the program was introduced in their county. The specifications are similar to the basic model, but they exclude the county time trends and the SS equation includes no MAF variable. As indicated in the notes to the table, the sample sizes are about a sixth as large as the full sample. Once again, the coefficients of the program variables are all positive and statistically significant. We note, however, that the MAF estimates are smaller than those in column 1, perhaps because some counties had not yet met all the quality standards imposed by the state. Finally, column 5 expands the basic sample by including all Hispanic births in the sample and attributing to the ones without test scores, the average test scores for Hispanics in the sample. Compared to the basic model, this modification has little effect on the estimated coefficients, which should alleviate concerns about the low matching rate for Hispanics shown in Table 1.[20]

Enriched Specifications

The enriched specifications replace the county-specific time trends with a variety of time varying variables at the county or school level, whose means and standard deviations are reported in the bottom three panels of Table 2. Our goal is to include as many potential confounders as possible so as to isolate the effects of the two programs. The new variables fall into three categories.

[20] We also estimated models to test the sensitivity of our results to two other issues. In one set, we added to the basic model a variable for whether a child was eligible for free and reduced-price lunch, imputing that variable and flagging the imputation as appropriate. The addition of that variable has little or no effect on the estimated program effects. Second, we reran the basic models using the smaller sample of matched births for which the county of birth and the county of third-grade schooling were the same. This sample is advantageous in that we can be quite confident about what county the child lived in during her early childhood years; it has the disadvantage that the movers to other counties may not be a random sample of the treated children. In any case, the program coefficients that emerge from this sample are very similar to those reported in the table.

The first category includes demographic variables at the county level by birth year. Included are the percentages of births to black mothers, to Hispanic mothers, to mothers with low education, as well as total births (in logarithmic form), all from the birth records. Other demographic variables are the percentages of population on the federal food stamp program and on the federal/state Medicaid program, both of which are oriented toward low-income households, as well as the total population and median family income. Birth cohorts with higher proportions of minority children or children from economically disadvantaged families are likely to exhibit lower performance on third-grade test scores than those with smaller proportions of disadvantaged children. Changes in the proportions of such children at the county level could plausibly be correlated with changes in funding for either or both of the other of the state's two early childhood programs, especially the MAF Program because it targets at-risk children. Median family income and total population could well be indicative of a county's capacity to implement either or both of the programs and further, the capacity and willingness of the county to support public education, or alternatively. In sum, there are conceptual reasons for including each of these variables as county-level control variables.[21]

A second category of variables are school funding variables measured during the year of the third-grade tests. These variables include per-pupil school funding (for all grades K–12, and possibly some preschool) by county (in 2009 dollars) financed by the federal, state, and local governments. In North Carolina, the state finances what it refers to as a "sound basic education," and local school districts have the option of augmenting state funding out of county tax revenues. As a result, the state is the main funder. More funding per pupil from state and local sources could potentially signify higher quality schooling, and could be associated with higher test scores. Alternatively, increases in funding from year to year in particular counties could reflect state or local policymaker's perceptions of the need for more funding because of low student performance. Federal funding for education is primarily for programs that target disadvantaged students so more funding of that type is likely to be associated with lower test scores. One danger with using information from the test score year is its potential endogeneity. To the extent that parents make choices among counties based on how their children are performing in school, the funding levels could be affected in part by the success of the early childhood programs.

The final measures are four characteristics of the school each child attended in third grade. They include two racial composition variables—the percentages of black students and students of other minority groups—and two measures of teacher characteristics—the percentage of inexperienced teachers and the average licensure test scores of all teachers in the school. Previous research shows that all four of these variables are typically associated with lower levels of student performance. Because these variables reflect decisions by parents, teacher labor markets, and school districts, all of which could well be influenced by the existence of the early childhood programs, they, like the funding variables, are not fully exogenous.

Table 4 presents models that successively add these categories of potential confounders to the basic model from which the county time trends have been eliminated. For convenience, we repeat estimates from the basic model in column 1. As shown in column 2, based on models that include all the county-level demographic variables, the program estimates remain positive and statistically significant. The magnitudes are smaller than those in column 1 by a third to a half. At the 2009

[21] In principle, it would have been desirable to include some other program-specific control variables, such as Head Start funding, but such data were not available at the county level during most of our time period. Hence, we are not able to draw any conclusions about the extent to which the state programs enhance or crowd out federal programs.

Table 4. Enriched models.

	Basic model (1)	Enriched model with county demographic variables (2)	Column 2 plus spending variables (3)	Column 3 plus school variables (4)
Panel A: Reading				
Smart Start ($00s)	0.0094***	0.0066***	0.0072***	0.0073***
	(0.0012)	(0.0010)	(0.0010)	(0.0012)
More at Four ($00s)	0.0187***	0.0119***	0.0187***	0.0195***
	(0.0032)	(0.0025)	(0.0032)	(0.0032)
Observations	886,783	886,783	886,783	824,376
Panel B: Math				
Smart Start ($00s)	0.0081***	0.0051***	0.0059***	0.0061***
	(0.0016)	(0.0013)	(0.0013)	(0.0014)
More at Four ($00s)	0.0214***	0.0097***	0.0163***	0.0182***
	(0.0040)	(0.0033)	(0.0040)	(0.0042)
Observations	890,827	890,827	890,827	827,880

Notes: Selected coefficients are from full regression models that include controls for a large number of child and parent characteristics at time of birth, child racial characteristics in grade 3, fixed effects for county, and year of birth. The full equations for columns 1 and 2 are included in the online Appendix, Table A1.[22] The basic model in column 1 includes county time trends. Column 2 replaces the county time trends with the county-level demographic variables in Table 2. Column 3 adds the per-pupil spending variables from Table 2, and column 4 add the school level variables from Table 2.The full equations for the basic model and the enriched model are included in the Appendix. Standard errors clustered at the county level are in parentheses.

***Statistical significance at the 0.01 level.

funding levels, the combined effects of the two programs in reading are about 0.2 standard deviations or about four months of instruction, and in math, about 0.16 standard deviations, or about two months of instruction. These are our preferred estimates of the true program effects.

Column 3 adds the school spending variables. With the addition of these variables, all four programmatic coefficients remain statistically significant. The same is true in column 4 where the models also include the four measures of school quality. Our interpretation of the patterns is that the preferred results in column 2 represent the effects of the two programs inclusive of the effects associated with the quality of a child's subsequent schooling, whereas those in the final column net out the quality of the schools and school systems after the early childhood years. With respect to the MAF Program, for example, the patterns suggest that the program itself raises the capacity of children to perform well in school, but some of that capacity is then diminished as children progress through public schools of low quality.[23]

[22] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.
[23] We remind the reader, however, that one should be careful about making causal statements with respect to the results in columns 3 and 4 because of the potential endogeneity of the spending and school quality variables that arise because they are measured during the testing year, not the birth year.

Robustness to Variation in Included Counties

Because our estimates are based on individual-level data and the sample sizes differ significantly across counties, an additional concern is that our results might be driven largely by the experience in particular counties and therefore not represent general patterns. For example, one or two large counties that might have more administrative capacity to implement the programs than more typical counties could potentially drive the results. To examine that possibility, we serially excluded each of the 10 largest counties from the analysis. In no case did the program effects change substantially or become nonsignificant, although they did decline somewhat with the removal of the largest counties. We report in Table A2 of the online Appendix the coefficients from the preferred model with the removal of each of the three largest counties: Charlotte Mecklenburg (CMS), Wake, and Guilford.[24] The results remain significant and are not substantially changed.

A related concern is that some of the early recipients of SS funding may have been more aggressive in making good use of the SS funding than some of the later recipients. Hence, including the initial pilot recipients in the sample might lead to upward biased estimates of what is possible in more typical counties. The final column of the online Appendix Table A2 provides information related to this issue. Contrary to our expectations, the sample that excludes the first 18 counties to receive SS funding generates somewhat larger average estimates of the effect of SS in both reading and math relative to the full sample. The explanation may well be that the initial counties were working in the context of a less well-developed state infrastructure and had to develop programs more on their own than was the case for counties that joined the program later and could learn from the experiences of others.

Effects by Subgroup

For several reasons, we hypothesize that the two programs would provide larger benefits for children from disadvantaged families than for those from more advantaged families. Although the SS Program does not specifically target children from economically disadvantaged households, we still expect such children to derive more educational benefits than their more advantaged counterparts whose families are in a better position to provide for them without state assistance. In particular, compared to their more advantaged peers, children from disadvantaged households may be more likely to participate directly in day care facilities or other activities funded with SS dollars, to benefit from easier access to the range of public services for families and children, and to benefit more once they are in elementary school from being in classrooms with children who are more prepared for school than they would have been without the program. The counterhypothesis is that an enrichment program such as SS will paradoxically widen the achievement gap if relatively advantaged children are better able to make use of its resources.[25] These hypotheses are tested against the null hypothesis that the program exerts equal benefits across groups. The case for differential benefits is even stronger for the MAF Program, which directly

[24] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

[25] An alternative hypothesis is that the state program may harm some disadvantaged students by crowding out other early childhood center such as Head Start. Such an outcome would be consistent with Cascio's (2009) finding for public subsidies for kindergarten. In contrast to her situation in which the public kindergarten was generally lower quality than the federal Head Start programs, the state's programs in this case are likely to be higher quality.

targets at-risk children. Benefits to relatively advantaged children would come from spillover effects alone.

We test for differential effects by adding a series of interaction terms to the initial models. In particular, we interact each of two program variables with an indicator for whether the child's mother has less than a high school education, is black, or is Hispanic. The results for the basic (trend) model and the enriched model with county covariates (column 2 of Table 4) are displayed in Table 5.[26] The coefficients for the first three variables in the table are the main effects for children with mothers who have low education, are black or Hispanic, and are all negative.

For each program and subject, the base coefficient of the program variable applies to a child whose mother has at least a high school degree and is neither black nor Hispanic. The positive coefficients for these program variables, all but one of which are statistically significant, consistently indicate that these advantaged families on average derive benefits from the two programs. These positive coefficients provide evidence of spillover effects, especially for MAF, which is targeted at at-risk children. The other coefficients in the table represent the additional program effects associated with the specified characteristics. Positive interaction effects with respect to mother's education emerge clearly and consistently for all three specifications. These positive coefficients indicate that children whose mothers have limited education benefit more from MAF than children with more educated mothers, with the differential effects being particularly large for the MAF Program.

We find little evidence of differentially positive or negative effects for children with black mothers from either program, after controlling for mother's education level.[27] For children with Hispanic mothers, the patterns are mixed. The evidence suggest that children of Hispanic mothers receive smaller (but still positive) benefits from the SS Program in reading than their peers, but somewhat larger benefits from the MAF Program in math. The smaller effect in reading may well reflect the more limited participation of Hispanics in preschool programs and quality child care (Hernandez, Denton, & Macartney, 2008; Liang, Fuller, & Singer, 2000).

CONCLUSION

This study of the population-level impacts of two early childhood programs in North Carolina on third-grade outcomes contributes to the literature in several ways. First, by measuring average impacts not on the program participants themselves, but rather on all children in a community who could be affected, the study provides a more comprehensive measure of the benefits of state investments in early childhood programs than do narrower studies that focus only on program participants. Second, by focusing on third-grade outcomes, this study sheds light on intermediate-run outcomes of potential interest to policymakers, and contrasts with recent studies

[26] The regression models on which these results differ from the basic regressions in the following two inconsequential ways. First, we replaced the original mother's education variables, which had been specified as a continuous variable with an indicator for 11 or fewer years of education to make the main effect comparable to the interaction variable. Second, we left out the black and Hispanic child indicator variables to simplify the interpretation of the variables for mother's race. We also reran the interaction models focusing on the child's race rather than the mother's race and obtained similar results.

[27] This finding deserves further research that is beyond the scope of this paper. It could be that the black children in fact derive initially larger benefits from the programs which then, as was found in the Head Start studies, dissipated as they attended schools with large concentrations of disadvantaged students and low-quality teachers. The net result could be an equal impact across groups. Alternately, it could be that the key differentiating variable is the education level of the mother and not the race of the mother or the child. In ongoing research we hope to sort out these two hypotheses.

of state-level programs based on regression discontinuity designs that were limited to examining short-run impacts or with other studies that have shown evidence of fadeout as children progress into formal schooling. Third, by examining two different, but complementary, early childhood programs, the study provides new insights into early childhood programs. One limitation of the study design, however, is that we are not able to say anything about what the effects of the MAF Program would have been in the absence of SS.

The findings are clearly positive. Each program generates average positive effects for reading and math scores in third grade. The magnitude of impact based on our preferred model (from column 2 in Table 4) and 2009 spending levels is equivalent to about two to four months of instruction in third grade. As we have illustrated in Tables 3 and 4, the estimates vary from specification to specification, with the lowest estimates emerging from models that do not account for any county-specific variation in county-wide variables over time. Our preferred specification controls for a large number of county-level demographic variables to rule out other factors that affect test scores that might be correlated with the programmatic variables of interest.

Our methodology does not allow us to parse the estimated total effects into those on program participants and those on others. Indeed, under the SS Initiative, the concept of a program participant is not obvious in many cases, given the multiple goals of the initiative and variation in the way it was implemented at the local level. For the MAF Program, the number of funded slots statewide at the program peak was about 20 percent of the full matched basic sample. If it were the case that all those funded slots were filled by children in the sample, and that there were no spillover effects of the program to other children, the test score impacts on the funded children would be unrealistically large: about five times the estimates reported in our tables. In other words, the coefficient of 0.0097 for math in column 2 of Table 5 implies that each $100 investment in MAF per four-year old would generate about a 0.05 average gain in math test scores for the funded students. Thus, an investment at 2009 funding levels of about $1,100 would imply gains of more than 0.55 standard deviations for funded students. Such estimates are far too large to be plausible, especially given the effect size of 0.10 to 0.14 for participants reported in the earlier study of the MAF Program (Peisner-Feinberg & Schaaf, 2010). The more plausible explanation for our large estimated impacts is that the program generates large positive spillovers to nonfunded children, which after all was a key goal of the program. Further support for the existence of positive spillovers comes from our investigation of effects by subgroup. Even for the MAF Program, which is targeted at low-income children, our estimates indicate that children of advantaged mothers also received benefits. Given that this group was ineligible for the program, the effects of direct funding must have spilled over.

In addition, we find that the effects of each program are somewhat larger for children whose mothers have low education than for those with more educated mothers, a finding that is consistent with other studies (e.g., Currie, 2001; Havenes & Mogstad, 2011). While it is clear that, on average, children of black mothers score far less well than children of white mothers on third-grade tests, both programs appear to be equally effective in raising the test scores of the two groups of children. The situation differs for children of Hispanic mothers, for whom we find a more mixed pattern of differential effects across programs and subjects, although the effects are still positive.

Currently, the state of North Carolina spends about $8,500 per child per year on elementary education. Based on a 10-month school year, the estimated effects based on funding levels of about $1,100 per child for each program translate into instruction costing between $1,700 and $3,400 per child in the population (not per directly treated child). Thus, based on effects on third-grade test score gains alone,

Table 5. Programmatic effects on test scores, with subgroup interactions, basic and enriched models.

	Basic (trend) model		Enriched model (preferred)	
	Reading	Math	Reading	Math
Low ed. mother	−0.0293***	−0.0110	−0.0285***	−0.0107
	(0.0084)	(0.0092)	(0.0084)	(0.0093)
Black mother	−0.248***	−0.302***	−0.246***	−0.297***
	(0.0226)	(0.0258)	(0.0230)	(0.0268)
Hispanic mother	0.0206	0.0207	0.0292	0.0294
	(0.0233)	(0.0273)	(0.0239)	(0.0282)
Smart Start (SS) $00s	0.0088***	0.0076***	0.0061***	0.0049***
	(0.0013)	(0.0018)	(0.0011)	(0.0015)
SS × low ed. mother	0.0029***	0.0023***	0.00290***	0.0023***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
SS × black mother	0.0001	1.23×10^{-5}	−0.0001	−0.0005
	(0.0009)	(0.0009)	(0.0010)	(0.0009)
SS × Hispanic mother	−0.0034*	−0.0026	−0.0043**	−0.0034
	(0.0020)	−0.0022	(0.0020)	(0.0023)
More at Four (MF) $00s	0.0168***	0.0170***	0.0010***	0.0053
	(0.0039)	(0.0041)	(0.0030)	(0.0034)
MF × low ed. mother	0.0042**	0.0067***	0.0039**	0.0067***
	(0.0017)	(0.0017)	(0.0018)	(0.0018)
MF × black mother	0.0017	0.0047	0.0020	0.0045
	(0.0027)	(0.0030)	(0.0027)	(0.0029)
MF × Hispanic mother	0.0007	0.0103**	0.0022	0.0102**
	(0.0045)	(0.0044)	(0.0045)	(0.0042)

***Statistical significance at the 0.01 level.

**Statistical significance at the 0.05 level.

*Statistical significance at the 0.10 level.

the state's investment of $2,200 per student for the early childhood programs would seem to be generating solid returns. In related research, we have documented that both programs also reduced the probability that children are identified as special needs students as of third grade, with somewhat varying effects, as would be expected by program and type of special need (Muschkin, Ladd, & Dodge, 2012). Such findings are consistent with the theoretically and empirically grounded prediction that early identification or intervention can reduce the subsequent need for special education services. If we were to add in the savings associated with the reduction in special education placements, the financial picture for the two programs would look even brighter. Although it is premature to do a full benefit-cost analysis at this time given that we have not compiled the data required to measure effects on outcomes beyond third grade, the results presented in this paper suggest that both programs would readily pass a benefit-cost test.

HELEN F. LADD is Edgar Thompson Professor of Public Policy and Professor of Economics, Sanford School of Public Policy, Duke University, Sanford Building 214A, Box 90245, Durham, NC 27708.

CLARA G. MUSCHKIN is Assistant Research Professor of Public Policy, Sanford School of Public Policy, Duke University, 220 Rubenstein Hall, Box 92000, Durham, NC 27708.

KENNETH A. DODGE is William McDougall Professor of Public Policy, Sanford School of Public Policy, Duke University, 222 Rubinstein Hall, Durham, NC 27798.

ACKNOWLEDGMENTS

This paper is part of an ongoing project by the authors evaluating early childhood programs in North Carolina. The authors are grateful to the Smith Richardson Foundation for its financial support to Duke University through the Beyond Test Scores Project. Helen Ladd also thanks the Center for the Analysis of Longitudinal Data in Education Research (CALDER) funded by the U.S. Department of Education and Kenneth Dodge acknowledges the National Institute of Drug Abuse Senior Scientist Award K05 DA015226. The authors thank Sarah Fuller and Richard Brady for their significant contributions to the analysis.

REFERENCES

Ananat, E. O., Gassman-Pines, A., & Gibson-Davis, C. M. (2011). The effects of local employment losses on children's educational achievement. In G. J. Duncan & R. Murnane (Eds.), Whither opportunity? Rising inequality and the uncertain life chances of low-income children (pp. 299–314). New York, NY: Russell Sage Foundation.

Baker, M., Gruber, J., & Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. Journal of Political Economy, 116, 709–745.

Barnett, W. S. (2011). Effectiveness of early educational intervention. Science, 333, 975–978.

Barnett, W. S., Epstein, D. J., Carolan, M. E., Fitzgerald, J., Ackerman, D. J., & Friedman, A. H. (2010). The state of preschool 2010: State preschool yearbook (executive summary). Rutgers, NJ: The National Institute for Early Education Research.

Besley, T., & Burgess, R. (2004). Can labor regulation hinder economic performance? Evidence from India. Quarterly Journal of Economics, 119, 91–134.

Bryant, D., Bernier, K., Peisner-Feinberg, E., & Maxwell, K. (2002). Smart Start and child care in North Carolina: Effects on quality and changes over time. Chapel Hill, NC: University of North Carolina, FPG Child Development Institute.

Bryant, D., Maxwell, K., Taylor, K., Poe, M., Peisner-Feinberg, E., & Bernier, K. (2003). Smart Start and preschool child care quality in North Carolina: Change over time and relation to children's readiness. Chapel Hill, NC: University of North Carolina, FPG Child Development Institute.

Cascio, E. (2009). Do investments in universal early education pay off? Long-term effects of introducing kindergartens into public schools. National Bureau of Economic Research, working paper 14951. Cambridge, MA: National Bureau of Economic Research.

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

Currie, J., & Thomas, D. (1995). Does Head Start make a difference? American Economic Review, 85, 341–364.

Davis-Kean, P. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. Journal of Family Psychology, 19, 294–304.

Duncan, G. D, & Magnuson, K. (2013). Investing in preschool programs. Journal of Economics Perspectives, 27, 109–132.

Fitzpatrick, M. D. (2008). Starting school at four: The effect of universal Pre-Kindergarten on children's academic achievement. The B. E. Journal of Economic Analysis & Policy, 8, 1–40.

Gormley, W. T., Jr. (2011). From science to policy in early childhood education. Science, 333, 978–981.

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. Developmental Psychology, 41, 872–884.

Havenes, T., & Mogstad, M. (2011). No child left behind: Subsidized child care and children's long run outcomes. American Economic Journal: Economic Policy, 3, 97–129.

Hernandez, D. J., Denton, N. A., & Macartney, S. E. (2008). Children in immigrant families: Looking to America's future. Social Policy Report, 22, 3–12.

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, 357–384.

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. Quarterly Journal of Economics, 122, 159–208.

Maxwell, K., Bryant, D., & Miller-Johnson, S. (1999). A six-county study of the effects of Smart Start child care on kindergarten entry skills. Chapel Hill, NC: The University of North Carolina, Frank Porter Graham Child Development Center.

Mervis, J. (2011). Giving children a head start is possible—But it's not easy. Science, 333, 956–957.

Muschkin, C., Ladd, H. F., & Dodge, K. A. (2012). Impact of North Carolina's early childhood initiatives on special education placements in third grade. Unpublished manuscript, Sanford School, Duke University, Durham, NC.

Peisner-Fineberg, E. S., Elander, K. C., Maris, C. L., & the More at Four Evaluation Team. (2006). Evaluation of the North Carolina More at Four Pre-Kindergarten Program, year 4. Chapel Hill, NC: The University of North Carolina, Frank Porter Graham Child Development Institute.

Peisner-Feinberg, E. S., & Schaaf, J. M. (2010). Long-term effects of the North Carolina More at Four Pre-Kindergarten Program: Children's reading and math skills at third grade. Chapel Hill, NC: Frank Porter Graham Child Development Institute.

Ponder, K. (2011). The Smart Start story. Unpublished manuscript.

Reynolds, A., Temple, J., Robertson, D., & Mann, E. (2002). Age 21 cost-benefit analysis of the Title I Chicago Child-Parent Centers. Educational Evaluation and Policy Analysis, I24, 267–303.

Wolfers, J. (2006). Did unilateral divorce laws raise divorce rates? A reconciliation and new results. American Economic Review, 96, 1802–1820.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. Journal of Policy Analysis and Management, 27, 122–154.

APPENDIX

Table A1. Coefficients for basic and enriched models.

	Basic model		Enriched model	
	Reading	Math	Reading	Math
Smart Start ($00s)	0.0094***	0.0081***	0.0066***	0.0052***
	(0.0012)	(0.0016)	(0.0010)	(0.0013)
More at Four ($00s)	0.0187***	0.0214***	0.0119***	0.0097***
	(0.0032)	(0.0040)	(0.0025)	(0.0033)
Extremely low birth weight	−0.273***	−0.479***	−0.272***	−0.479***
	(0.0137)	(0.0138)	(0.0138)	(0.0140)
Very low birth weight	−0.138***	−0.249***	−0.139***	−0.250***
	(0.0112)	(0.0117)	(0.0112)	(0.0116)
Low birth weight	−0.0866***	−0.132***	−0.0866***	−0.132***
	(0.0034)	(0.0038)	(0.0034)	(0.0039)
High birth weight	0.0468***	0.0585***	0.0467***	0.0585***
	(0.0032)	(0.0031)	(0.0032)	(0.0031)
Female	0.182***	0.00305	0.182***	0.0029
	(0.0034)	(0.0034)	(0.0034)	(0.0034)
Mother's education (years)	0.110***	0.113***	0.110***	0.113***
	(0.0015)	(0.0016)	(0.0015)	(0.0016)
Marital status	0.0878***	0.0901***	0.0880***	0.0904***
	(0.0053)	(0.0047)	(0.0054)	(0.0048)
Mother's age (years)	0.0112***	0.00710***	0.0113***	0.0072***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
No dad information	−0.0507***	−0.0519***	−0.0499***	−0.0507***
	(0.0060)	(0.0057)	(0.0063)	(0.0059)
Mother immigrant	0.0735***	0.132***	0.0741***	0.134***
	(0.0141)	(0.0135)	(0.0142)	(0.0136)
First born	0.212***	0.115***	0.212***	0.115***
	(0.0038)	(0.0030)	(0.0038)	(0.0030)
Mother black	−0.245***	−0.297***	−0.245***	−0.297***
	(0.0202)	(0.0242)	(0.0202)	(0.0242)
Mother Native American	−0.159***	−0.164***	−0.160***	−0.165***
	(0.0243)	(0.0201)	(0.0246)	(0.0201)
Mother Asian	0.0769***	0.161***	0.0761***	0.160***
	(0.0212)	(0.0269)	(0.0213)	(0.0274)
Mother Hispanic	−0.0207	0.0074	−0.0212	0.0055
	(0.0145)	(0.0158)	(0.0148)	(0.0159)
Mother other race	0.116***	0.0906*	0.113***	0.0861*
	(0.0355)	(0.0488)	(0.0366)	(0.0488)
Child black	−0.296***	−0.332***	−0.295***	−0.331***
	(0.0118)	(0.0132)	(0.0118)	(0.0135)
Child Native American	−0.225***	−0.212***	−0.225***	−0.212***
	(0.0159)	(0.0159)	(0.0158)	(0.0160)
Child Asian	−0.0945***	−0.0037	−0.0953***	−0.0033
	(0.0307)	(0.0305)	(0.0311)	(0.0308)
Child Hispanic	−0.199***	−0.122***	−0.199***	−0.120***
	(0.0122)	(0.0153)	(0.0122)	(0.0153)
Child mixed race	−0.112***	−0.163***	−0.111***	−0.161***
	(0.0098)	(0.0101)	(0.0097)	(0.0100)
Constant	−1.823***	−1.533***	−4.521***	−4.113*
	(0.0233)	(0.0257)	(1.305)	(2.245)
County and year fixed effects?	Yes	Yes	Yes	Yes
County time trends?	Yes	Yes	No	No
County demographic variables?	No	No	Yes	Yes
Observations	886,783	890,827	886,783	890,827
R-squared	0.237	0.247	0.236	0.245

***Statistical significance at the 0.01 level.
*Statistical significance at the 0.10 level.

Table A2. Estimates based on samples with some counties excluded.

	Full sample	Without CMS	Without wake	Without Guilford	Without SS pilot counties
Smart Start ($00s)					
Reading	0.0066***	0.0058***	0.0066***	0.0068***	0.0067***
	(0.0010)	(0.0008)	(0.0011)	(0.0010)	(0.0010)
Math	0.0051***	0.0042***	0.0055***	0.0056***	0.0057***
	(0.0013)	(0.0010)	(0.0013)	(0.0013)	(0.0011)
No. of observations					
Reading	886,783	814,990	822,564	840,874	691,981
Math	890,827	818,620	826,428	844,715	695,071
More at Four ($00s)					
Reading	0.0119***	0.0110***	0.0109***	0.0111***	0.0106***
	(0.0025)	(0.0023)	(0.0023)	(0.0024)	(0.0025)
Math	0.0097***	0.0087***	0.0082***	0.0087***	0.0084***
	(0.0033)	(0.0031)	(0.0030)	(0.0031)	(0.0035)
No. of observations					
Reading	886,783	814,990	822,564	840,874	691,981
Math	890,827	818,620	826,428	844,715	695,071

Notes: Selected coefficients from regression models similar to the enriched regression model in column 2 of Table 5. See notes to that table.

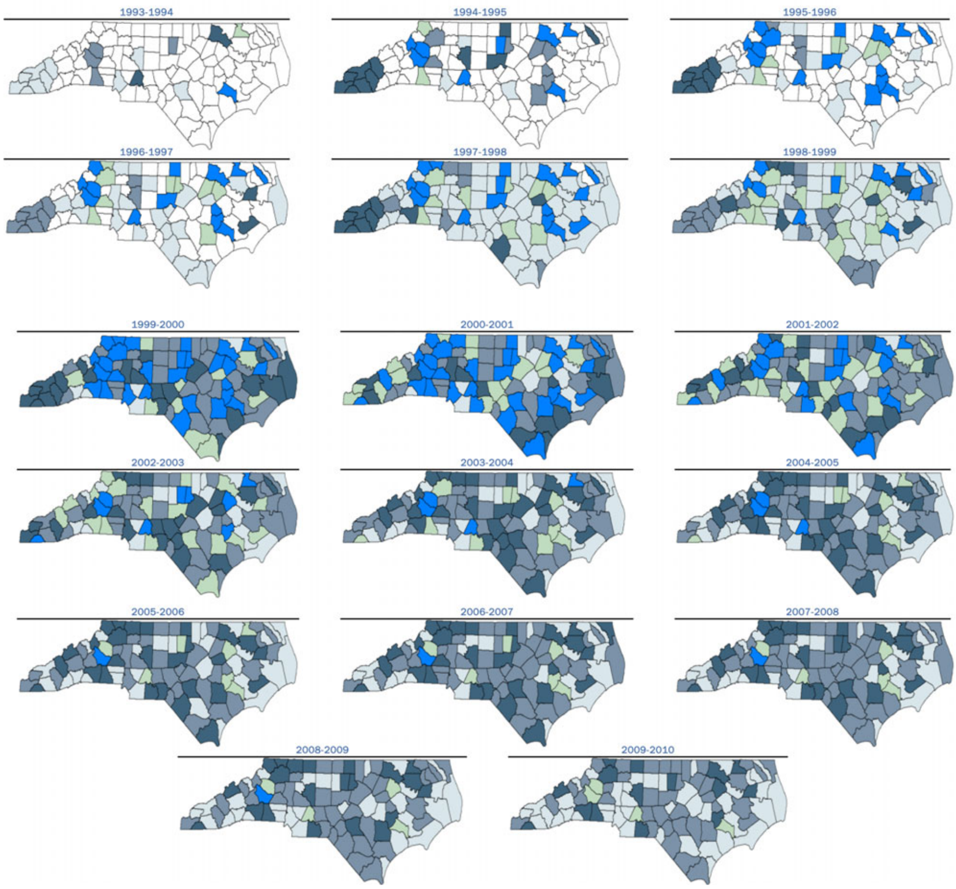
***Statistical significance at the 0.01 level.

CMS denotes Charlotte-Mecklenburg School System.

STATEWIDE FUNDING FIGURES
FY 1993-2009
10/6/2013

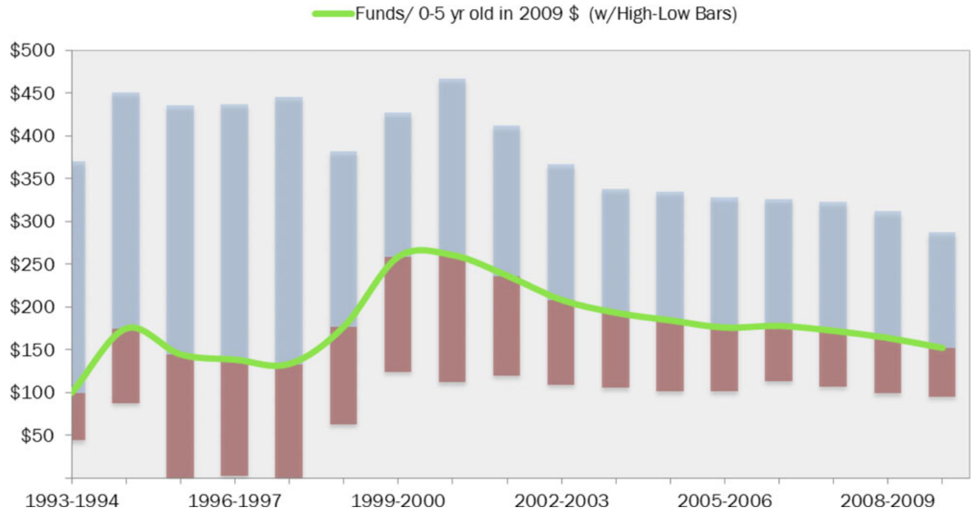
COUNTY-LEVEL PENETRATION TREND: REAL FUNDS RATIO

VARIATION AMONG COUNTIES	1993-1994	1994-1995	1995-1996	1996-1997	1997-1998	1998-1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010
Real Funds Ratio Range	# of Counties served in each range																
< $150	10	3	13	22	56	48	7	11	5	18	23	26	30	25	26	33	42
$150 - $200	4	8	3	9	6	19	32	29	37	34	36	36	35	39	43	41	39
$200 - $250	2	11	7	2	10	7	22	13	17	21	27	28	27	29	25	21	15
$250 - $300	1	2	6	9	9	16	10	16	22	18	10	7	7	6	5	4	4
$300+	1	8	15	13	17	10	29	31	19	9	4	3	1	1	1	1	-
Not participating	82	68	56	45	2	-	-	-	-	-	-	-	-	-	-	-	-



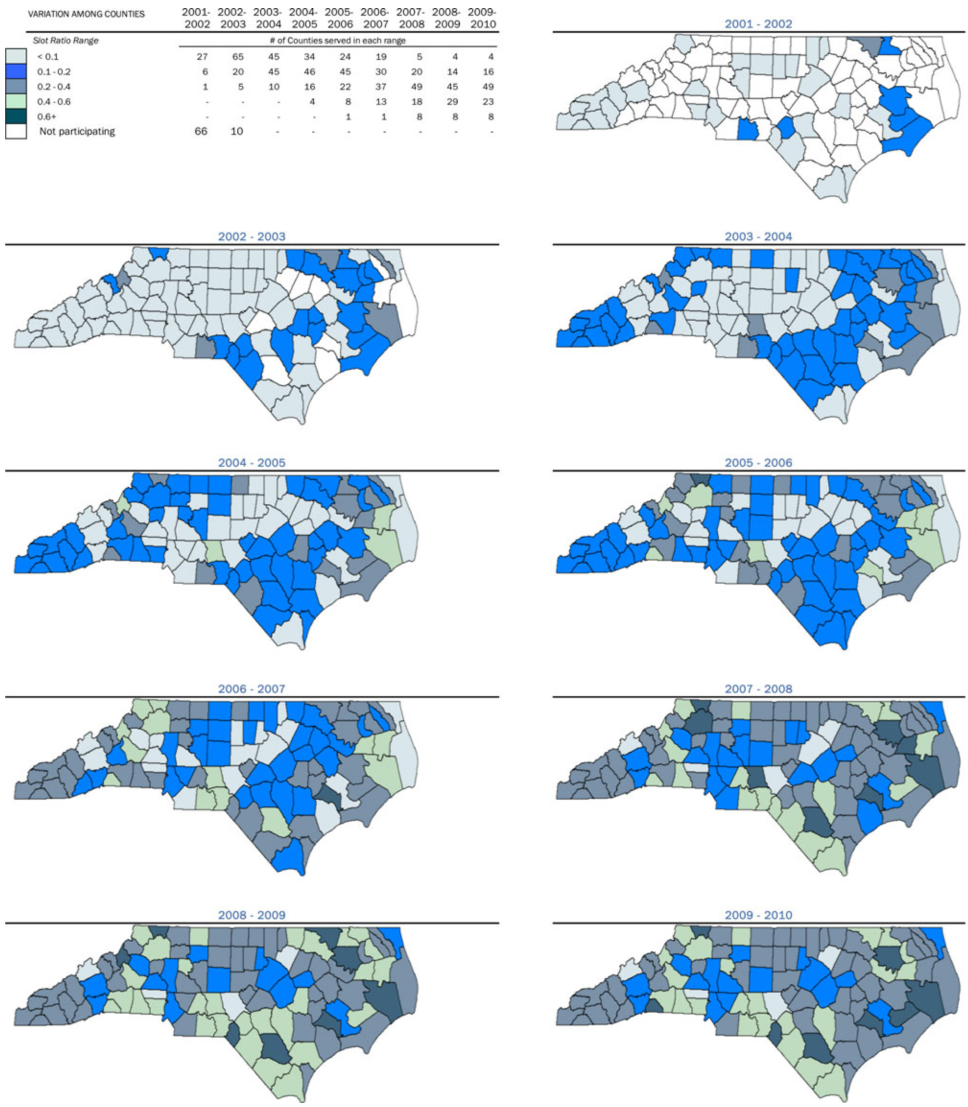
Notes:
1. *Data Sources*
   (a) Yearly Smart Start Funding data provided by North Carolina Partnership for Children FY 1998 to 2009, NC Division of Child Development FY 1993 to 1997.
   (b) Yearly, county-level zero- to five-year-old population estimates provided by North Carolina Office of State Budget and Management.
   (c) Monthly CPI data provided by Bureau of Labor Statistics, U.S. Department of Labor.
2. County Real Funds Ratio—Smart Start operating funds allocations deflated to July 1993 dollars divided by total estimated zero- to five-year-old population.
3. Real Funds—All data classified as “Real Funds” are in July 1993 dollars using the CPI as of July in each year as a deflator.

Figure A1. Geographic Distribution of Smart Start Funding over Time.



Notes:
1. *Data Sources*
   (a) Yearly Smart Start Funding data provided by North Carolina Partnership for Children FY 1998 to 2009, NC Division of Child Development FY 1993 to 1997.
   (b) Yearly, county-level zero- to five-year-old population estimates provided by North Carolina Office of State Budget and Management.
   (c) Monthly CPI data provided by the, Bureau of Labor Statistics, U.S. Department of Labor.
2. Penetration Measures
   (a) County Nominal Funds Ratio—Smart Start operating funds allocations divided by total estimated zero- to five-year-old population
   (b) County Real Funds Ratio—Smart Start operating funds allocations deflated to July 1993 dollars divided by total estimated zero- to five-year-old population
3. Real Funds—All data classified as Real Funds are in July 1993 dollars using the CPI as of July in each year as a deflator.

Figure A2. Smart Start Penetration Rates.



VARIATION AMONG COUNTIES	2001-2002	2002-2003	2003-2004	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010
Slot Ratio Range	# of Counties served in each range								
< 0.1	27	65	45	34	24	19	5	4	4
0.1 - 0.2	6	20	45	46	45	30	20	14	16
0.2 - 0.4	1	5	10	16	22	37	49	45	49
0.4 - 0.6	-	-	-	4	8	13	18	29	23
0.6+	-	-	-	-	1	1	8	8	8
Not participating	66	10	-	-	-	-	-	-	-

Figure A3. Geographic Distribution of More at Four Spending over Time.

Notes:
1. *Data Sources*
   (a) Yearly More at Four Funding and Slot Allocation data provided by North Carolina Office of Early Learning.
   (b) Yearly, county-level four-year-old population estimates provided by North Carolina Office of State Budget and Management.
   (c) Monthly CPI data provided by Bureau of Labor Statistics, U.S. Department of Labor.
2. Penetration Measures
   (a) County Slot Ratio—Number of allocated More at Four slots divided by total estimated four-year-old population.
   (b) County Nominal Funds Ratio—More at Four operating funds allocations divided by total estimated four-year-old population.
   (c) County Real Funds Ratio—More at Four operating funds allocations deflated to July 2001 dollars divided by total estimated four-year-old population
3. Real Funds—All data classified as Real Funds are in July 2001 dollars using the CPI as of July in each year as a deflator.
4. 2001 to 2004 Funding Allocations—A proportion of the funding in any contractor's initial year is used as "start-up expenses"; hence, the $/slot figures may be slightly overstated in the first few years of the program.

Figure A4. More at Four: Slots and State Funding.

Notes:

1. *Data Sources*
   (a) Yearly More at Four Funding and Slot Allocation data provided by North Carolina Office of Early Learning.
   (b) Yearly, county-level four-year-old population estimates provided by North Carolina Office of State Budget and Management.
   (c) Monthly CPI data provided by Bureau of Labor Statistics, U.S. Department of Labor.
2. Penetration Measures
   (a) County Slot Ratio—Number of allocated More at Four slots divided by total estimated four-year-old population.
   (b) County Nominal Funds Ratio—More at Four operating funds allocations divided by total estimated four-year-old population.
   (c) County Real Funds Ratio—More at Four operating funds allocations deflated to July 2001 dollars divided by total estimated four-year-old population.
3. Real Funds—All data classified as Real Funds are in July 2001 dollars using the CPI as of July in each year as a deflator.
4. 2001 to 2004 Funding Allocations—A proportion of the funding in any contractor's initial year is used as start-up expenses; hence, the $/slot figures may be slightly overstated in the first few years of the program.

Figure A5. More at Four: County-Level Penetration.