

Impact of North Carolina's Early Childhood Programs and Policies on Educational Outcomes in Elementary School

Kenneth A. Dodge, Yu Bai, Helen F. Ladd, and Clara G. Muschkin
Duke University

North Carolina's Smart Start and More at Four (MAF) early childhood programs were evaluated through the end of elementary school (age 11) by estimating the impact of state funding allocations to programs in each of 100 counties across 13 consecutive years on outcomes for all children in each county-year group ($n = 1,004,571$; 49% female; 61% non-Latinx White, 30% African American, 4% Latinx, 5% other). Student-level regression models with county and year fixed effects indicated significant positive impacts of each program on reading and math test scores and reductions in special education and grade retention in each grade. Effect sizes grew or held steady across years. Positive effects held for both high- and low-poverty families, suggesting spillover of effects to nonparticipating peers.

Surging policy interest in early education and care programs for children from birth to age 5 has heightened demand for rigorous evaluation of programs and policies to determine which strategies produce the largest impacts on promoting child development. Because of inconsistent findings across studies (to be described below), policy and scientific interest has focused on whether positive effects of early childhood programs endure or fade. "Fadeout" refers to diminishing differences between an intervention and control group over time following an initial positive impact of a program due to either decrement in performance by the intervention group or deferred improvement by the control group. The current study examines whether initial positive impacts of North Carolina's early childhood programs are sustained or fadeout by the end of elementary school.

The field has moved from exclusive reliance on the findings of a handful of small experiments begun in the 1960s to examination of the impact of policies and programs as they are implemented at scale. This progression follows from the Institute of Medicine's prescription for bringing basic science to policy by moving from scientifically based small randomized trials of efficacy in the laboratory to

larger, more naturalistic trials in the community and, finally, to evaluation of programs when implemented at scale (Mrazek & Haggerty, 1994). Coincident with the shift to evaluation of programs at scale is growing interest in whether state-level policies and programs can "move the needle" to improve population-level outcomes by exerting impact not only on program participants but also on their peers through a spillover effect, and whether these impacts are sustained or fadeout by the end of elementary school.

The current study evaluates fadeout and spillover of impact of North Carolina's two flagship early childhood programs on education outcomes for 13 cohorts of about one million children as they develop from birth through the end of elementary school. This study is the third in a series using similar methods and data. Ladd, Muschkin, and Dodge (2014) found initial positive impact of each program on reading and math test scores in Grade 3, and Muschkin, Ladd, and Dodge (2015) found impact on reducing special education placements in Grade 3. The current study contributes to the literature in two important ways: First, we estimate impact on children as they develop through the end of elementary school across Grades 3, 4, and 5, to test whether early positive impact fades out, is sustained, or grows. Second, we estimate "spillover" of impact on the classmates of children who had been

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.

Correspondence concerning this article should be addressed to Kenneth A. Dodge, Center for Child and Family Policy, Duke Box 90545, Durham, NC 27708-0545. Electronic mail may be sent to dodge@duke.edu.

© 2016 The Authors
Child Development © 2016 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2017/8803-0021
DOI: 10.1111/cdev.12645

served. Both sets of analyses inform models of how early interventions affect children's development.

Intervention Basis in Developmental Science

Early childhood programs and policies are based in twin tenets contributed by developmental science. The first is that long-term outcomes for individuals are produced to an inordinate degree by environmental experiences in the first several years of life, which set children on trajectories that persist across development (Shonkoff, 2014). The second tenet follows that achievement gaps across ethnic and income groups begin during these early childhood years and then persist as children get older. Two theories dominate explanations for the persistence, or even growth, of achievement gaps: a family investment model and a stress-induced family process model (Yeung, Linver, & Brooks-Gunn, 2002). The family investment model posits that higher income enables some families to invest more in resources for children. Although low-income families have increased their investment over the past several decades, the differences across income levels in family investment have actually widened (McLanahan & Jacobsen, 2015). The stress-induced family process model asserts that the stress of financial disadvantage causes parents to be distracted from their normal parenting behavior or to act in harsh or less effective ways. Both models suggest that children in disadvantaged homes experience relatively less cognitive stimulation (e.g., books, meaningful conversations, Hart & Risley, 1995) during critical years of synaptic development and that communities in which large portions of children come to kindergarten from early disadvantage tend to have schools characterized by in-classroom remediation, high rates of grade retention and special education, and lower academic achievement for the entire population. The challenges of learning in the contexts of high disadvantage may be exacerbated by spillover effects that occur through peer influences, interrupted teaching, and lower teaching levels (Davoudzadeh, McTernan, & Grimm, 2015).

These perspectives suggest two intervention strategies that have been developed to address achievement gaps (Haskins & Rouse, 2005) and now characterize the great majority of government policies and program options for early education and child care (Dodge & Haskins, 2015). The first strategy is high-quality, early out-of-home education and child care, coupled with family support and a holistic approach to academic, social, and

health outcomes. Government has supported high-quality child-care and preschool education models (e.g., Head Start, Perry Preschool, Abecedarian, Child Parent Centers) for disadvantaged children, improvement in a community's infrastructure of early education and child care through rating and licensing of child-care facilities and professional development of child-care workers, and direct subsidies to low-income families so that they can seek equitable child care in the marketplace. This kind of intervention grew rapidly at the turn of the 21st century, but government support has waned since 2008 (Barnett, Carolan, Squires, & Clarke Brown, 2013).

The second intervention strategy is direct enrollment in high-quality academic preschool during the year immediately prior to kindergarten, usually targeted toward children meeting income needs-based criteria, but in some cases implemented universally. Pre-kindergarten (pre-K) intervention focuses less on ameliorating parenting problems or preventing early cognitive deprivation and instead attends to building curricular-based reading and math skills just prior to the beginning of formal schooling. From a population perspective, this strategy is intended to improve achievement trajectories of low-income children, close achievement gaps, and boost achievement for all children by populating kindergarten classrooms with more children ready to learn. Funding for pre-K programs grew in the first half of the first decade of this century but waned slightly with the economic downturn (Barnett, Carolan, et al., 2013). North Carolina developed prototypic national models for early child care and pre-K, called Smart Start (SS) and More at Four (MAF), respectively, which are the focus of this report.

Past Evaluations of Early Childhood Programs

Evaluation of several small early programs fueled excitement in the field. A randomized trial of the Perry Preschool Program (Schweinhart et al., 2005a, 2005b) that provided home visiting and enriched out-of-home child care at ages 3 and 4 showed favorable impact on cognitive and language development, but these gains faded out during elementary school only to re-emerge in high school graduation, employment, income, and well-being. A randomized trial of the Abecedarian Project (Campbell & Ramey, 1994) that provided home visiting and out-of-home child care from infancy to age 5 showed positive impact on cognitive and language development, partial fading during

elementary school, and re-emergent long-term effects in reading and math achievement and educational attainment. A matched comparison-group evaluation of the Chicago Child Parent Centers (Temple & Reynolds, 2007) that provided home visiting and enriched preschool indicated positive impact on test scores at kindergarten entry, but effects subsequently faded, although were still significant through eighth grade and into adulthood.

Although these programs provide evidence for the potential positive impacts of early education and care, the samples were small and are not representative of most preschool and child-care programs delivered by government at scale, which are typically less expensive per child, less intense, and less carefully monitored. Most evaluations of scaled-up programs have involved nonexperimental designs. Gormley and Gaye (2005) compared consecutive cohorts and found that Tulsa's universal pre-K program increased test scores by a third of a standard deviation after 1 year. Fitzpatrick (2008) found that the Georgia program increased math and reading scores, with effects lasting through fourth grade. The New Jersey Abbott Preschool Program Longitudinal Effects Study (Barnett, Jung, Youn, & Frede, 2013) found significant impact on reading and math test scores and grade retention through Grades 4 and 5.

Andrews, Jargowsky, and Kuhne (2012) found that participation in the Texas pre-K program was associated with about a 0.06 *SD* increase in third-grade math and reading scores and reductions in grade retention and special education services. The longest follow-up of state pre-K is a study of Michigan's Great Start School Readiness program, which found initial test score gains that faded out in middle school but re-emerged on high school exit examinations (Schweinhart, Xiang, Daniel-Echols, Browning, & Wakabayashi, 2012).

Two randomized trials of scaled-up programs yielded similar findings of initial positive impact but complete fadeout of effects within several years. The Head Start Impact Study showed that random assignment to the program was associated with short-term gains in cognitive and social skills that faded out during elementary school (U.S. Department of Health and Human Services, Administration for Children and Families, 2010). A randomized trial of Tennessee's pre-K program indicated that assignment to the program was associated with initial positive impact on cognitive skills, complete fadeout of effects 1 year later, and adverse impact after several years (Lipsey, Farran, & Hofer, 2015).

A meta-analysis by Camilli, Vargas, Ryan, and Barnett (2010) and a systematic narrative review by Yoshikawa et al. (2013) lead to the general conclusion that early childhood and pre-K programs implemented at scale through policy have, on average, initial positive effects on program participants that partially fade throughout elementary school to about half the size of the initial effects. These impacts vary with program quality, depending on the subsequent learning environment. We know little about the very long-term effects of most early childhood programs.

Developmental Models of Fadeout, Persistence, and Spillover

Even a high-quality early childhood experience that provides a child with strong kindergarten readiness is not sufficient to lead to sustained positive learning outcomes. Early childhood programs do not operate as "immunizations" against subsequent environmental adversity; rather, development occurs as an ongoing transaction between the child and her or his learning environment (Sameroff, 2009). Bailey, Duncan, Odgers, and Yu (2015) propose three mechanisms through which initial positive impact of an early childhood program could be sustained across development: skills building, foot in the door, and sustaining environments. If the conditions that allow these mechanisms are in place, then one would hypothesize that early impacts should be maintained across development, whereas if they are not in place, then one would anticipate fadeout of early impact before the end of elementary school. The first mechanism of sufficient skills building depends on the early childhood program being of high quality. The second mechanism is that a program might have immediate impacts at a sensitive period in life that indirectly cause other impacts that cascade into long-term outcomes. Dodge, Pettit, and Bates (1994) proposed and found support for a developmental cascade in the etiology of adolescent substance use, starting with early childhood parenting problems that led to a child's failure at school, which led to the child's gravitation toward deviant peers, and so on. A similar cascade but toward positive proximal and distal outcomes might be initiated by an early childhood intervention. The third mechanism is based on the assumption of an early childhood intervention for disadvantaged children: The hypothesis is asserted that if a child from any background comes to kindergarten with sufficient skills, then a high-quality school learning environment populated by other

children who are ready to learn will sustain that child's educational development through the end of elementary school. A corollary is that if the school environment is of poor quality, then early gains will not be sustained. Thus, we assert that variation in long-term impacts of early childhood programs can be explained at least partially by the degree to which the subsequent peer and learning environment supports and sustains early skill development.

"Sustaining environments" in elementary school could mean a peer environment indexed by the number of peers who had also received early enrichment. Justice, Petscher, Schatschneider, and Mashburn (2011) found that within-year growth in preschool children's language abilities is influenced by the prior language abilities of classroom peers, and Hanushek, Kain, Markman, and Rivkin (2003) found that third graders' math scores are influenced by the average prior score of a child's peers. Using a matched-peer study design, Curenton, Dong, and Shen (2015) found that a child's gains made in pre-K are sustained through fifth grade only to the extent that the child's peers' achievement is high.

Because peer effects are reciprocal, children who benefit from enriched early childhood experiences could also influence peers positively, called spillover. Spillover of early program impact to nonparticipant peers is more likely when a high proportion of classmates had also participated in an early childhood program. Thus, we assert that both persistence of early impact on participants and spillover to nonparticipant peers depend on the penetration level of early childhood programs; that is, implementation that occurs at scale or at least with a high proportion of children participating will have a higher level of sustained impacts and less fadeout.

North Carolina's two early flagship childhood programs offer an intriguing case because each was designed to be implemented at full scale and thus could have sustained effects on participants and spillover to nonparticipant peers. Naturally occurring cross-time and cross-community variation in the penetration of these programs provide the comparative basis for testing the hypothesis that sustenance (vs. fadeout) of population impacts across the period of elementary school will depend on the degree of penetration of these early childhood programs.

North Carolina's Two Flagship Early Childhood Programs

We evaluate the impacts of an early childhood program and a pre-K program in North Carolina

that have been rated as meeting high-quality standards and have been rolled out in a way that resembles a natural experiment to enable greater rigor in evaluation.

The SS Initiative

Smart Start, a prototype of the first intervention strategy type, is best conceptualized and measured as state dollars allocated to counties to improve early childhood services for children from birth to age 4 to ensure that all children enter school healthy and ready to learn. Services were intended to benefit all children; although in practice, disadvantaged children were most directly targeted because they were eligible for child-care subsidies. SS began as a series of demonstration projects in a small number of counties and was later expanded across years to serve all counties. This variation in spending across counties and years constitutes the independent variable in the current study.

The MAF Pre-K Program

More at Four (since called NC pre-K) is the state-funded pre-K program for high-risk 4-year-old children, with risk defined as annual family income at or below 75% of the state median, limited English proficiency, disability, chronic illness, or developmental need. MAF was initiated in 2001 to improve school readiness among these children in the year prior to kindergarten matriculation.

State funds were provided to qualifying pre-K programs for each qualifying child enrolled in that classroom. Each funded classroom had to meet state-defined standards for staff credentials, class size, teacher-child ratios, and state child-care licensing requirements. Most slots were filled in classrooms also populated by nonfunded children, so MAF compounded its reach by requiring the entire classroom to meet standards. Past evaluators have found that classroom quality standards were maintained throughout the process of scaling-up MAF (N.C. Office of State Budget and Management, 2014; Peisner-Feinberg; the More at Four Evaluation Team, 2008) so that across the years covered by this evaluation NC's pre-K programs consistently reached the highest possible rating of 10 for benchmark standards for quality as defined by the National Institute of Early Education Research (Barnett, Carolan, Squires, Clarke Brown, & Horowitz, 2015). At peak funding, it reached most eligible 4-year-olds and many more peers with a program rated as high-quality; thus, we hypothesized that

this high penetration rate would lead many later kindergarten classrooms to provide a sustaining environment for any early impacts.

Evaluation Design and Research Questions

We avoided problems of selection bias into treatment and unmeasured spillover of impact to nearby nontreated children by conceptualizing an entire age cohort of children within a community as receiving the same level of intervention service, which we operationalized as a continuous dimension of state dollars allocated per child and to compare children across communities and across age cohorts. We treated differences in dollar allocations as exogenous to any particular child because it was highly unlikely that a family would move to a different community in anticipation of next year's county allocation. Thus, we could treat the state's allocation scheme as a natural experiment at the county-year level. State allocations for SS to a county in a year were largely uncorrelated with allocations for MAF, and the impact of SS was uncorrelated with impact of MAF, adding credence to this natural experiment assumption.

In earlier manuscripts, we evaluated the impact of SS and MAF on children's outcomes at Grade 3. Ladd et al. (2014) found positive impact of each program on third-grade reading and math test scores, and Muschkin et al. (2015) found positive impact of each program on lowering special education placements through Grade 3. Although these findings are sound, this research has not yet followed children into the age when fadeout of program effects has been documented in other states. In the current article, we extended the analysis to examine impacts through the end of elementary school across Grades 3, 4, and 5, and we included outcomes of math and reading test scores, special education placements, and grade retention. We refrained from strong hypotheses and used two-tailed tests to answer specific questions.

The first question is whether program impacts change as children get older. Past studies of fadeout in other early childhood programs suggest that the impact of North Carolina's programs might decline as children experience continued neighborhood, family, and school disadvantage. Even more worrisome is the possibility that initial positive impact on reduced special education placement and grade retention might simply delay an inevitable downward trajectory into school failure or, even worse, exacerbate this failure by deferring children from receiving necessary placements in early grades. In

contrast, developmental studies of how school and peer environments reinforce and embellish the impact of early positive experiences (Bradshaw, Waasdorp, & Leaf, 2012), along with Heckman's (Cunha & Heckman, 2007) concept that skill begets skill, suggest that gains due to early interventions might actually grow across development. One rather unique feature of North Carolina's implementation is that it attempted to saturate a community with its early childhood programs so that a child would enter the kindergarten classroom surrounded by peers who are ready to learn and a teacher who would not have to remediate others, thus possibly leading to maintenance of early gains.

The second question is whether program impacts varied significantly across subgroups within the population, defined by maternal education, family income, and child ethnicity. Previous analyses of Grade 3 outcomes indicated that positive impacts of both SS and MAF held for each subgroup, but the magnitude of effects on reading and math scores and special education placement rates was greater for children of mothers with low education (Ladd et al., 2014; Muschkin et al., 2015). Because the programs (especially MAF) emphasized penetration with the low-income subgroup, an obvious hypothesis is that program impacts would be greater for this group; however, possible spillover effects suggest that impacts might extend to all groups within a county cohort. Of particular interest in the current study is whether early impacts of programs would be sustained across development to a greater degree for certain subgroups. It might well be that disadvantaged children benefit immediately from an early childhood program, but their impacts fade out to a greater degree than for other children because continuing disadvantage at home provides a nonsustaining learning environment. Therefore, it is crucial to examine subgroup effects across development.

Method

Research Participants

The state's vital records provided birth information and the NC Department of Public Instruction provided school enrollments through Duke University's North Carolina Education Research Data Center. Participants included every child born in North Carolina (NC) between January 1, 1988 and December 31, 2000, who attended a public school in NC at least some time between July 1, 1995 and June 30, 2012, whether or not the child actually participated

in a state-funded early childhood program. Of the 1,347,562 births during this period, 1,004,571 were matched as attending a public or charter school at some time in Grades 3, 4, or 5 (74.55% of children born in NC during this period). This match rate is higher than the 68% previously reported by Ladd et al. (2014) and Muschkin et al. (2015) because additional years of education files were scrutinized that allowed for re-entry of children who had been missing in Grade 3 but came back in Grades 4 or 5, corrections were made to clerical problems with matching, and newer cohorts were added for whom state data quality improved. Nonmatches occurred because of migration out of state, attendance at a private school, or errors in the matching process.

We compared birth record characteristics of matched and unmatched children. Detailed variable definitions are listed in the appendix tables of Ladd et al. (2014) and Muschkin et al. (2015). Birth records were coded for the child's gender, race, and ethnicity (coded as White, Black, Asian, Native American, Hispanic, or other), status as firstborn or not, and birth weight (scored as extremely low [$< 1,000$ g], very low [$\geq 1,000$ and $1,500$ g], low [$\geq 1,500$ g and $< 2,500$ g] or high [$\geq 2,500$ g]); mother's age, marital status (coded as single or married at time of birth), education level (in years), race and ethnicity, and immigrant status; and whether the father is identified. The matched group is more likely than the unmatched group to have a mother who is single, a nonimmigrant, and non-White. These differences are likely due to the fact that unmatched children attend private schools or have Spanish surnames that were mistakenly recorded in school files. All birth-record variables were included as covariates in analyses.

Independent Variables

The design tested impact of county-year state allocations for SS and MAF on individual-level educational outcomes for the entire population of children in NC across the study period. Available outcome variables in administrative records included standardized reading and math test scores in Grades 3, 4, and 5, and special education placements and grade retentions through the end of fifth grade. To improve precision of measurement and to account even further for any possible selection bias, we incorporated a variety of individual-level variables available from birth records and time-varying county-level variables as covariates. Controlling time-varying county-level variables is important to fend off the possibility that annual variations in a

county's funding allocations might be correlated with a county's year-specific needs. Incorporating county fixed effects in the analysis further controls stable county differences, and time fixed effects controls secular changes in state-level economic conditions. Exposures to SS and MAF were scored from administrative records of state funding allocations for each program to each county by year. All funding levels were calibrated to 2009 dollars. Of the 13 birth cohorts from 1988 to 2000, 8 cohorts experienced only SS (1988–1995) and 5 cohorts experienced both SS and MAF (1996–2000).

Smart Start

Smart Start is not a manualized program but, rather, a series of locally administered state funding supports for birth to age 5 child care and education through day care vouchers, health and family support services, child-care services, professional development, and promoting collaborations among local agencies. Bryant, Bernier, Peisner-Feinberg, and Maxwell (2002) found that counties participating in SS did indeed improve their average classroom quality, average teacher qualifications, and the proportion of child-care centers with national accreditation and licensure.

Although 30% of all local SS dollars were required to be spent on child-care services for qualifying low-income children, program developers anticipated that impact would spread to all children who attended child-care centers because of improved standards and, eventually, to all same-cohort kindergarteners who would enroll in classrooms that were filled by peers who had directly benefitted from SS. For primary analyses, we mapped per-child dollar allocations for each of the 100 counties for each of the 13 birth cohorts of children across the study period and matched them to specific children based on their county of residence at birth.

SS began in 1993 in 18 counties, increased to more than 50 counties by 1997 and was implemented in all 100 NC counties by the 1998–1999 school year. The state selected initial counties for implementation somewhat arbitrarily but to represent each congressional district. Funding grew from zero to a peak of \$250 million (in 2009 \$) in 2000 at an average per child aged 0 to 4 of about \$400 per child per year and fell to about \$220 per child per year in 2009 (the last year included in the current evaluation). Because a child could benefit from funds for 5 years across ages 0–4, the total state funds allocated per child within a county for

children in the current study averaged \$1,100 ($\220×5 years) and ranged from \$0 to \$3,500, with high variability within and between counties over the study period (Ladd et al., 2014).

More at Four

More at Four funding was provided to counties to support classroom-based slots for eligible children not for specific preschool programs. Across the study period, the state provided an average of about \$4,400 per slot for 1 year and reached up to 25% of all 4-year-olds by 2010. Like SS, funding began in pilot counties, and varied in the number of counties and dollars per student across years (Ladd et al., 2014). About half of MAF slots were in public school settings, a third in for-profit community child-care centers, and the remainder in nonprofit child-care centers and Head Start programs. The average MAF investment was about \$330 per 4-year-old (countywide average, even though only about one-fourth of students were enrolled); however, this average cost includes the years before MAF was initiated in 2001; for the years when MAF was funded fully, the MAF investment per 4-year-old in the population was about \$1,100, similar to the SS investment. Therefore, we used \$1,100 per child in the population for SS and MAF when we describe the effect sizes of the state's investment; note that all analyses use the continuous investment variable in dollars ranging from zero upward. We defined "treatment" by the level of MAF funding (per age-appropriate child) in the county in a year.

Dependent Variables and Covariates

Four outcome variables come from educational administrative data files available at the North Carolina Education Research Data Center in each of the 3 years corresponding to a child's expected grade levels at Grades 3, 4, and 5. If a child was not found within that cohort, other grade levels were examined in case the child was enrolled in one or two grade levels above or below the expected level. The terms "Grade 3," "Grade 4," and "Grade 5" are applied for school years that corresponded to the expected grades for a child based on age but could in fact be a higher or lower grade level. For each of Grades 3, 4, and 5, reading and math achievement scores were taken from the end-of-grade-standardized tests administered by the state in May of each school year. Because scoring criteria changed across years, scores were standardized ($M = 0$, $SD = 1$)

within each grade and each school year based on our matched sample. *Special education placement* was based on the exceptionality information contained in the end of year data files submitted by the Accountability and Exceptional Children divisions of the NC Department of Public Instruction. A child was coded as being placed in special education for disability if any nongifted class of exceptionality was noted. As reported by Muschkin et al. (2015), specific learning disabilities and speech-language impairments were the most frequent subcategories coded. The percent of students with a special education placement was 15.00% in Grade 3, 15.70% in Grade 4, and 15.60% in Grade 5. *Grade retention* was identified if a child was found in the same grade in consecutive years. The percent of children who repeated Grade 3 was 2.85%, the percent who repeated Grade 4 was 1.24%, and the percent who repeated Grade 5 was 0.91%. We also coded a variable for "ever repeated a grade," which had a base rate of 4.69%.

Covariates drawn from birth records and education files include 10 child characteristics, 11 mother characteristics, and 8 county-year variables. Child covariates include 0–1 indicators for extremely low birth weight, very low birth weight, low birth weight, and high birth weight (with normal birth weight as reference group in each case); ethnicity with Black, Native American, Asian, Hispanic, and mixed race groups (with White as reference group in each case); and residential move (1 if birth count and school county are the same, 0 otherwise). Mother characteristics include education (in years), marital status (1 for married, 0 for not married), age (in years), father information (1 for present, 0 for missing), immigrant status (1 for yes, 0 for not), primipara (1 for yes, 0 for no), and ethnicity with Black, Native American, Asian, Hispanic, and mixed race groups (with White as reference group in each case). Time-varying county-level covariates include percent births to Black mothers, percent births to Hispanic mothers, percent births to low-education mothers, percent population receiving food stamps, percent population receiving Medicaid, number of births (log), total population (log), and median family income (in 2009 \$). Free or reduced lunch (FRL) status (coded as 0 = no, 1 = yes, or missing) is the indicator of poverty and was drawn from the school records because information about poverty at birth was not available. The proportion of missing cases for this variable for Grades 3, 4, and 5, were 15.0%, 7.53%, and 2.94%, respectively. Because we had a child's scores for this variable in every school

year, we used the mode of a child's nonmissing values to impute the score, which reduced missing values to 0.34%, 0.18%, and 0.13% for Grades 3, 4, and 5, respectively. All analyses were conducted with this variable included as a covariate as well as without this variable (which slightly increased the sample size), without substantive differences in findings.

Data Analyses

Of the children for whom birth records and school records were matched, 21.20% moved from one NC county to another between birth and Grade 3. Group means and tests of differences between movers and nonmovers for available variables can be found online at www.childandfamilypolicy.duke.edu. Movers were more likely than nonmovers to be White or Hispanic, firstborn, have a younger mother, and have a mother with lower education.

Because the data files did not contain records of the child's county of residence between birth and elementary school, we repeated multiple regression models with different assumptions. We conducted analyses with the entire sample, presuming the child was exposed to SS and MAF in the county that the child resided at birth. As a robustness check, we also ran models with the subgroup of children who resided in the same county at birth and school, the subgroup that moved between birth and Grade 3 (with the SS and MAF variables assigned according to the county of birth), and with the subgroup that moved (with SS and MAF assigned according to the county of residence at Grade 3).

We tested impact of SS and MAF in two sets of analyses, the first testing impact at a single grade (3, 4, or 5) and the second testing impact on growth in the outcome over time using a hierarchical linear model (HLM). The single-grade basic model takes the following form,

$$O_{icbtg} = \beta_1 SS_{ict}^* + \beta_2 MAF_{ict}^* + \beta_3 X_{itb} + \beta_4 Y_{it} + \beta_5 C_{cb} + Grade_{ig} + \alpha_c + \gamma_b + \epsilon_{icbt},$$

where O_{icbtg} is a Grade g outcome ($g = 3, 4,$ and 5 ; such as math and reading score, grade repeating, grade repeating ever, special education status, and special education status ever) in Year t for the i th student born in county c in Year b . All models included county fixed effects (α_c) and year

fixed effects (γ_b) in order to control for any overall county differences and secular trends in outcomes. The analyses effectively become tests of differences in SS and MAF funding within counties across time, netting out any statewide time trends. Linear regression models were used when the dependent variables were reading and math scores. Logistic regression models were used when the dependent variables were grade repeated (0/1), grade repeat ever (0/1), special education status (0/1), and special education status ever (0/1).

All models included fixed covariates at the student level (X , i.e., child gender; birth weight as extremely low, very low, low, or high; race as Black, Hispanic, Asian, Native American, or other; mother's education level; marital status; age; father information available on birth record; immigration status; firstborn status; and race), and time-varying covariates at the county level (C , i.e., proportion births to Black mothers, proportion births to Hispanic mothers, proportion births to low education mothers, number of births, total population, median family income, population with food stamps, and population with Medicaid), in order to control for any factors that might correlate with funding allocations. Because only 0.7% of observations had missing values, listwise deletion was used in all models.

Hierarchical linear models of within-child growth curves took the following form:

Level 1: repeated measures within student,

$$O_{gic} = \pi_{0ic} + \pi_{1ic}(\text{Time Point})_{gic} + \epsilon_{gic}$$

where, O_{gic} is the outcome at Grade g for child i who born in county c ; $(\text{Time Point})_{gic}$ is 0 at Grade 3, 1 at Grade 4, and 2 at Grade 5; π_{0ic} is the initial status of child i_c , that is, the expected outcome for that child in the Grade 3 (when Time Point = 0); and π_{1ic} is the learning rate for child i_c during academic year; and ϵ_{gic} is an error assumed to be normally distributed with a mean of 0 and a variance σ^2 .

Level 2: student level,

$$\begin{aligned} \pi_{0ic} &= \beta_{00c} + \beta_{01c}X_{ic} + r_{0ic}, \\ \pi_{1ic} &= \beta_{10c} + r_{1ic}, \end{aligned}$$

β_{00c} is the mean initial status within county c where child i was born; β_{10c} is the mean growth rate of the outcome variable within county c where child i

was born; X_{ib} is a set of variables of student characteristics.

Level 3: county level,

$$\beta_{00ct} = \gamma_{000t} + \gamma_{001t}SS_{ict} + \gamma_{002t}MAF_{ict} + \gamma_{003t}C_c + \mu_{00ct},$$

$$\beta_{10ct} = \gamma_{100t} + \gamma_{101t}SS_{ict} + \gamma_{102t}MAF_{ict} + \mu_{10ct}.$$

γ_{000t} is the overall mean initial status; γ_{100t} is the overall mean growth rate; γ_{001t} is the effect of SS on initial status (i.e., score); γ_{002t} is the effect of MAF on initial status (i.e., score); γ_{101t} is the effect of SS on learning rate; γ_{102t} is the effect of MAF on learning rate. C_c is a set of county-level variables at County c where child i was born.

Results

Descriptive Analysis

Table 1 describes the sample: 906,914 in the Grade 3 panel, 910,744 in the Grade 4 panel, and 910,527 in the Grade 5 panel. Eighty-two percent of students were born with normal birth weight. The majority of students were non-Hispanic White, 30.20% of students were non-Hispanic Black, 3.64% were Hispanic students, 1.01% were Asians, and 1.81% were Native Americans. The distribution of race groups for children was similar to that for their birth mothers. The average years of education for birth mothers were 12.53, and their average age was 25.85 years. Sixty-six percent of children were born to mothers who were married at the time of their birth. Forty-four percent had no siblings at birth.

Preliminary Analyses

As expected, most birth-record covariates significantly predicted outcomes, as can be found in Appendices S2 and S3. Individual-level covariates for race and ethnicity of child and parent, low birth weight, low mother education, single-parent status, young mother age, mother status as immigrant, and noninvolvement of father were robust predictors of low math and reading scores at each grade, placement in special education at each grade and ever, and grade retention at each grade and ever. Few community-level covariates predicted outcomes after individual-level variables were controlled. All covariates were included in all models.

We estimated HLM for math and reading test scores to test whether significant variation was

present at each of three levels (within children across years, across children, and across counties). Variation in math scores indicated that Level 1 factors (within child across years) accounted for 27.81% of the total variance in scores, Level 2 factors (between child) accounted for 59.27% of total variance, and Level 3 factors (county) accounted for 12.93% of total variance. The HLM for reading scores indicated that Level 1 factors (within child across years) accounted for 29.11% of the total variance in scores, Level 2 factors (between child) accounted for 58.21% of total variance, and Level 3 factors (county) accounted for 12.68% of total variance. We retained the three-level model for analyses.

We report analyses in three stages: First, we tested model coefficients for each program at each single point in time (Grades 3, 4, and 5) and coefficients for growth curve analyses that test whether program impact grows or fades. Although we have reported some Grade 3 analyses in prior articles, by adding records for Grades 4 and 5 we capture a slightly increased sample size for growth curve models and here report new Grade 3 findings, which are very similar to those reported previously. Second, we tested moderation of program effects across groups. Third, we examined whether program coefficients are robust to various assumptions made about county of residence at time of SS and MAF program delivery.

Impact of SS and MAF on Student Outcomes at Each Single Time Point

Reading and Math Test Scores

The upper panel of Table 2 lists model coefficients for the impact of SS and MAF funding allocations on standardized math test scores in Grades 3, 4, and 5. Grade 3 coefficients have been reported previously (Ladd et al., 2014), but are included for comparison purposes. Minor variations in coefficients from those reported previously are due to the inclusion of a higher proportion of children in Grades 4 and 5. Significant positive coefficients indicate that state SS allocations to a child's county in a year for which that child was eligible to be served by those allocations were robustly associated with higher math scores for that child when she or he was in Grades 3 ($B_{SS} = .0051$, $p < .01$), Grade 4 ($B_{SS} = .0050$, $p < .01$), and Grade 5 ($B_{SS} = .0056$, $p < .01$). When math scores in Grade 3 were added as a covariate in the models, SS was a marginally significant

Table 1
 Descriptive Characteristics of Research Participants

Variables	Grade 3 (n = 906,194)		Grade 4 (n = 910,744)		Grade 5 (n = 910,527)	
	M	SD	M	SD	M	SD
Academic outcomes						
Math standardized test score	0.00	1.00	0.00	1.00	0.00	1.00
Reading standardized test score	0.00	1.00	0.00	1.00	0.00	1.00
Grade retention (%)	2.85	16.60	1.24	11.10	0.91	9.48
Grade retention ever (%)	—	—	3.92	19.40	4.69	21.10
Special education placement (%)	15.00	35.70	15.70	36.40	15.60	36.30
Special education placement ever (%)	—	—	17.70	38.20	19.00	39.20
Program						
Smart Start allocations per child over 5 years (\$00's)—only non-zero allocations	11.51	8.60	11.46	8.61	11.41	8.61
More at Four allocations per 4-year-old child (\$00's)—only non-zero allocations	3.34	2.49	3.32	2.48	3.31	2.48
Smart Start allocations per child over 5 years (\$00's)—all participants	9.86	8.92	9.81	8.92	9.74	8.92
More at Four allocations per 4-year-old child (\$00's)—all participants	1.03	2.07	1.01	2.05	0.99	2.04
Student characteristics (%)						
Female	49.10	50.00	49.10	50.00	49.10	50.00
Extremely low birth weight	0.46	6.77	0.46	6.77	0.46	6.74
Very low birth weight	0.80	8.93	0.81	8.93	0.81	8.95
Low birth weight	6.94	25.40	6.95	25.40	6.94	25.40
Normal birth weight	81.90	38.50	81.90	38.60	81.90	38.50
High birth weight	9.94	29.90	9.94	29.90	9.94	29.90
Child White	60.90	48.80	60.80	48.80	60.60	48.80
Child Black	30.20	45.90	30.20	45.90	30.30	46.00
Child Native American	1.81	13.40	1.83	13.40	1.85	13.50
Child Asian	1.01	10.00	1.01	10.00	0.99	9.91
Child Hispanic	3.64	18.70	3.63	18.70	3.67	18.80
Child mixed race	2.46	15.50	2.51	15.70	2.56	15.80
Mother characteristics						
Mother's education (years)	12.53	2.42	12.53	2.41	12.53	2.41
Marital status (%)	66.40	47.20	66.40	47.20	66.30	47.20
Mother's age (years)	25.85	5.89	25.84	5.89	25.83	5.88
No dad information (%)	14.50	35.20	14.50	35.30	14.60	35.30
Mother immigrant (%)	5.97	23.70	5.92	23.60	5.86	23.50
Firstborn (%)	44.00	49.60	44.00	49.60	44.00	49.60
Mother White (%)	63.40	48.20	63.40	48.20	63.30	48.20
Mother Black (%)	30.10	45.90	30.20	45.90	30.20	45.90
Mother Native American (%)	1.65	12.80	1.66	12.80	1.68	12.80
Mother Asian (%)	1.15	10.70	1.14	10.60	1.13	10.60
Mother Hispanic (%)	3.62	18.70	3.58	18.60	3.54	18.50
Mother other race (%)	0.05	2.33	0.06	2.35	0.05	2.34
County-level demographic data by birth year						
Births to Black mothers (share of births), %	30.10	16.60	30.40	16.70	30.40	16.70
Births to Hispanic mothers (share of births), %	3.63	4.11	3.59	4.06	3.54	4.01
Births to low education mothers (share of births), %	23.40	5.77	23.60	5.80	23.50	5.77
Population on food stamps (share of population), %	7.45	3.76	7.45	3.76	7.45	3.76
Population on Medicaid (share of population), %	13.40	5.65	13.40	5.65	13.40	5.65
Number of births (log)	7.075	1	7.092	0.996	7.085	0.995
Total population (log)	11.71	0.988	11.71	0.987	11.71	0.986
Median family income (2009 \$)	\$57,699	\$10,451	\$57,669	\$10,436	\$57,639	\$10,417

Table 1
Continued

Variables	Grade 3 (n = 906,194)		Grade 4 (n = 910,744)		Grade 5 (n = 910,527)	
	M	SD	M	SD	M	SD
School characteristics, test year						
Black students (share of students), %	30.05	24.20	29.79	24.14	29.34	24.08
Other minority students (share of students), %	11.87	12.70	12.68	12.99	13.71	13.35
Per-pupil spending by source, test year						
Federal (2009 \$)	\$638	\$338	\$722	\$392	\$802	\$423
State (2009 \$)	\$4,172	\$1,050	\$4,074	\$1,178	\$3,961	\$1,300
Local (2009 \$)	\$2,307	\$1,422	\$2,623	\$1,614	\$2,939	\$1,783
Same county (%)	78.80	40.90	78.00	41.40	77.30	41.90

Table 2
Linear Regression Model Results for Math Scores and Reading Scores

	(1) Grade 3	(2) Grade 4	(3) Grade 4, Control Grade 3	(4) Grade 5	(5) Grade 5, Control Grades 3 and 4
Math					
Math at G3			.7622 (−.0024)***		.3251 (−.0018)***
Math at G4					.5438 (−.0016)***
Smart Start (\$00's)	.0051 (−.0013)***	.0050 (−.0012)***	.0010 (−.0006)*	.0056 (−.0013)***	−.0003 (−.0007)
More at Four (\$00's)	.0117 (−.0035)***	.0170 (−.0033)***	.0048 (−.0016)***	.0204 (−.0036)***	.0036 (−.0021)*
Observations	870,893	867,419	842,978	835,177	794,380
R ²	.247	.25	.682	.251	.745
Reading					
Reading at G3			.7400 (−.0026)***		.3407 (−.0018)***
Reading at G4					.4985 (−.0017)***
Smart Start (\$00's)	.0069 (−.0011)***	.0062 (−.0011)***	.0010 (−.0005)*	.0077 (−.0012)***	.0009 (−.0005)*
More at Four (\$00's)	.0144 (−.0026)***	.0198 (−.0031)***	.0059 (−.0015)***	.0224 (−.0039)***	.0037 (−.0021)*
Observations	867,116	862,815	837,666	831,155	789,058
R ²	.237	.247	.657	.248	.705

Note. Models include county and time fixed effects. Robust standard errors are in parentheses. All covariates presented in Table 1 are controlled in all models, but their coefficients are not listed. * $p < .1$. *** $p < .01$.

predictor of math scores in Grade 4 ($B_{SS} = .0010$, $p < .10$). When Grades 3 and 4 scores were covaried, SS did not predict Grade 5 math scores ($B_{MAF} = -.0003$, *ns*).

MAF allocations were significantly associated with higher math scores in Grade 3 ($B_{MAF} = .0117$, $p < .01$), Grade 4 ($B_{MAF} = .0170$, $p < .01$), and Grade 5 ($B_{MAF} = .0204$, $p < .01$), with coefficients increasing across years. To test for additional impact on a trajectory of development beyond the initial impact on Grade 3 scores, we ran models controlling for Grade 3 scores, and MAF was a significant predictor of math scores in Grade 4 ($B_{MAF} = .0048$, $p < .01$). When Grades 3 and 4 scores were

controlled, MAF did not significantly predict Grade 5 math scores ($B_{MAF} = .0036$, $p < .10$).

The lower panel of Table 2 lists model coefficients for program impacts on standardized reading test scores in Grades 3, 4, and 5. SS allocations were associated with significantly higher reading test scores in Grade 3 ($B_{SS} = .0069$, $p < .01$), Grade 4 ($B_{SS} = .0062$, $p < .01$), and Grade 5 ($B_{SS} = .0077$, $p < .01$), with consistent coefficients across years. When reading scores in previous years were covaried in the models, SS was still a statistically significant predictor of reading scores in Grade 4 ($B_{SS} = .0010$, $p < .1$) and Grade 5 ($B_{MAF} = .0009$, $p < .1$).

MAF allocations were associated with higher reading scores in Grade 3 ($B_{MAF} = .0144, p < .01$), Grade 4 ($B_{MAF} = .0198, p < .01$), and Grade 5 ($B_{MAF} = .0224, p < .01$), with coefficients increasing across years. When prior year scores were added as a covariate, MAF remained a statistically significant predictor of reading scores in Grade 4 ($B_{MAF} = .0059, p < .01$) but not for Grade 5 ($B_{MAF} = .0037, p < .10$).

To interpret the magnitudes of these coefficients, we compared children who live in a county with the average funding level for SS (\$1,100, calculated as \$220 per 0- to 5-year-old child per each of 5 years of eligibility of funding) with children who live in a county with no funding. Based on the coefficients reported above, an average investment by SS generated an average gain in Grade 4 math scores of 0.0550 SD (i.e., the reported coefficient of SS, $.0050 \times 11 = .055$) and in Grade 4 reading scores of 0.068 SD (i.e., the reported coefficient of SS, $.0062 \times 11 = .068$). We also compared children who resided in a county that received an MAF investment of \$1,100 per 4-year-old (the steady-state funding level once the program was implemented in all counties during this time period) with children who received no MAF funding. An \$1,100 MAF investment was associated with an average gain of 0.1870 SD in Grade 4 math scores ($.0170 \times 11$) and an average gain of 0.2178 SD

($.0198 \times 11$) in Grade 4 reading scores. The typical annual gain in test scores from the beginning to the end of Grade 4 is about 0.50 SD in reading and 0.75 SD in math (depending on the year, the math gain ranges from .6 to .9). These estimates are based on our own calculations as well as information reported in the North Carolina State Testing Results, 2008–2009.

If we assume a 10-month school year, an effect size of 0.1 SD represents about 2 months of instruction in reading and about 1.5 months in math. Figure 1 shows that the average early investment in MAF is associated with gains in both math and reading test scores and that the gains increase between Grade 3 and Grade 4 and between Grade 4 and Grade 5. The same figure shows that gains associated with average SS funding are also significantly greater than zero and only modestly increase across grades. The cumulative number of months gained in reading from SS and MAF combined was 4.1 for third graders, 5.1 for fourth graders, and 6.2 for fifth graders. The cumulative number of months gained in math from SS and MAF combined was 2.4 for third graders, 2.6 for fourth graders, and 3.3 for fifth graders.

The analyses above do not test differences in effects across grades. Next, we estimated a growth curve model with the multiple-time average test score as the intercept outcome and linear trend of

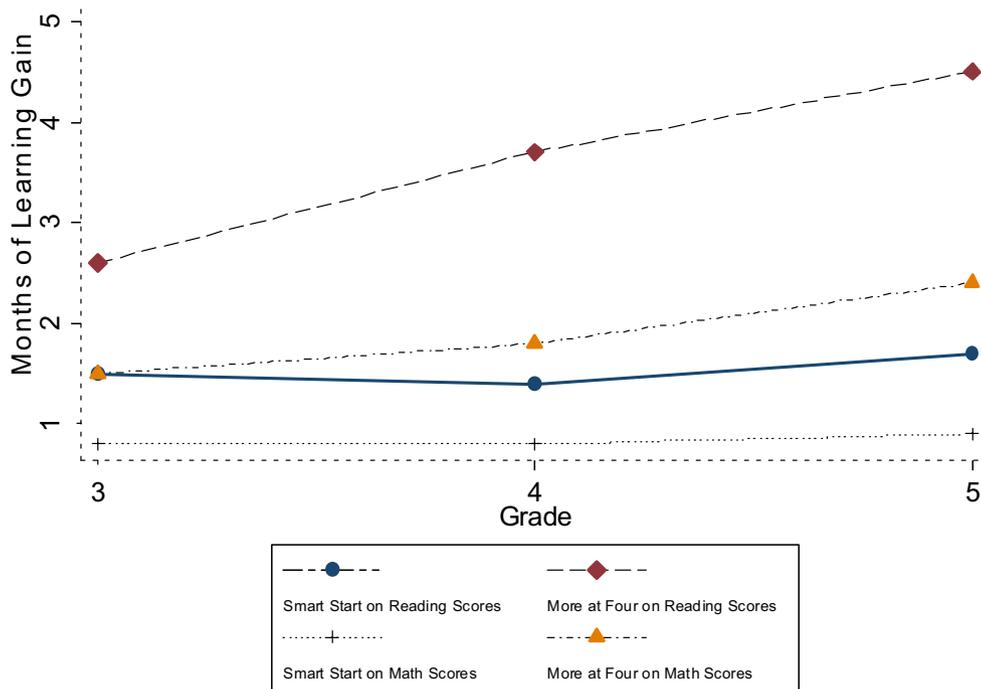


Figure 1. Months of reading and math learning gained by average allocations to Smart Start and More at Four versus no allocations.

scores across Grades 3 through 5 as the slope outcome. These models indicated that SS had a significant effect on math scores ($B_{SS} = .0017, p < .01$) and reading scores ($B_{SS} = .0032, p < .01$) averaged over years; and MAF had a significant effect on math scores ($B_{MAF} = .0069, p < .01$) and reading scores ($B_{MAF} = .0086, p < .01$) averaged across years. SS also had a significant effect on growth in math scores across grades ($B_{Grade \times SS} = .0004, p < .01$) but not reading scores ($B_{Grade \times MAF} = .0001, ns$). MAF did not have a significant effect on growth in math scores ($B_{Grade \times MAF} = -.0002, ns$) or reading scores ($B_{Grade \times MAF} = .0001, ns$).

Special Education Placement

The upper panel of Table 3 lists the odds ratios for program impacts on placement in special education in Grades 3, 4, and 5. The odds of being placed in special education in Grade 3 were significantly decreased by higher SS investments ($OR_{SS} = .9908, p < .05$). For those children born in a county with an average SS investment of \$1,100 per child versus no funding, the odds ratio was .90, $\exp(\ln(.9908) \times 11) = .90$, indicating that the average allocation of SS funding reduced the probability of being placed in special education by 10%. In Grade 4, the significant SS odds ratio ($OR_{SS} = .9922, p < .05$) indicated that the average investment reduced the probability of special education placement by 8%. In Grade 5, the significant odds ratio

($OR_{SS} = .9938, p < .05$) indicated that the average SS investment reduced the probability of special education placement by 7%.

Unit investments in MAF had a significant effect on the probability of being placed in special education in Grade 3 ($OR_{SS} = .9695, p < .001$), Grade 4 ($OR_{SS} = .9499, p < .01$), and Grade 5 ($OR_{SS} = .9415, p < .001$). The odds ratio of being placed in special education for a student born in a county with the average MAF investment of \$1,100 per child was .71 in Grade 3, $\exp(\ln(.9695) \times 11)$, .57 in Grade 4, $\exp(\ln(.9499) \times 11)$, and .52 in Grade 5, $\exp(\ln(.9415) \times 11)$, indicating that average MAF funding reduced the probability of special education placement by 29% in Grade 3, 43% in Grade 4, and 48% in Grade 5. These findings indicate that average SS funding allocations had a significant but modest effect of reducing special education placements by about 10% in each of Grades 3, 4, and 5. MAF had a significant and growing effect on reducing special education placements by up to 48% in Grade 5.

Grade Retention

Table 3 lists model coefficients for program impacts on repeating Grades 3, 4 (i.e., repeating Grade 4 no matter whether any previous grade had been repeated), and 5. SS allocations reduced the probability of repeating Grade 3 ($OR_{SS} = .9890, p < .01$). The odds ratio for repeating Grade 3 for students who were born in a county with the

Table 3
Logistic Regression Models on Special Education and Grade Retention

	Grade 3	Grade 4	Grade 5
1. Special Ed			
Smart Start (\$00's)	0.9908 (−.0025)***	0.9922 (−.0026)***	0.9938 (−.0024)**
More at Four (\$00's)	0.9695 (−.0068)***	0.9499 (−.0067)***	0.9415 (−.0069)***
Observations	897,342	899,928	897,547
2. Special Ed ever			
Smart Start (\$00's)	0.9908 (−.0025)***	0.9924 (−.0024)***	0.9933 (−.0023)***
More at Four (\$00's)	0.9695 (−.0068)***	0.9524 (−.0065)***	0.9476 (−.006)***
Observations	897,342	899,928	897,547
3. Grade repeat			
Smart Start (\$00's)	0.9890 (−.0062)*	0.9982 (−.0079)	0.9862 (−.012)
More at Four (\$00's)	1.004 (−.0125)	1.0071 (−.0207)	0.9972 (−.0282)
Observations	897,342	899,928	897,547
4. Grade repeat ever			
Smart Start (\$00's)	0.9890 (−.0062)*	0.9852 (−.0062)**	0.9848 (−.0068)**
More at Four (\$00's)	1.004 (−.0125)	0.9249 (−.0179)***	0.8993 (−.0203)***
Observations	897,342	899,928	897,547

Note. Models include county and time fixed effects. Robust standard errors are in parentheses. All covariates presented in Table 1 are controlled in all models, but their odds ratios are not listed. * $p < .1$. ** $p < .05$. *** $p < .01$.

average SS investment of \$1,100 per child was $.86$, $\exp(\ln(.9890) \times 11) = .89$, indicating that SS program reduced the probability of repeating Grade 3 by 11%. SS did not have a significant impact on new grade repetitions in Grade 4 ($OR_{SS} = .9982$, *ns*) or Grade 5 ($OR_{SS} = .9862$, *ns*). A new variable was created to indicate "grade retention ever" to indicate grade retention ever in Grades 3, 4, or 5. SS allocations reduced the probability of grade retention ever by Grade 5 ($OR_{SS} = .9848$, $p < .001$), and the odds ratio to repeat a grade ever was $.84$, $\exp(\ln(.9848) \times 11)$.

MAF allocations did not have a significant impact on repeating Grade 3 ($OR_{SS} = 1.0040$, *ns*), Grade 4 ($OR_{SS} = 1.0071$, *ns*), or Grade 5 ($OR_{SS} = .9972$, *ns*). However, MAF allocations had a significant impact on reducing the probability of ever repeating a grade by Grade 4 ($OR_{SS} = .9249$, $p < .001$) and Grade 5 ($OR_{SS} = .8993$, $p < .001$), and the odds ratio to repeat a grade ever was $.42$ by Grade 4, $\exp(\ln(.9249) \times 11)$, and $.31$ by Grade 5, $\exp(\ln(.8993) \times 11)$.

Moderation of Impact Across Poverty Subgroups

To test whether SS and MAF allocations had different effects on poverty subgroups, we repeated models with the addition of interaction terms (for $SS \times FRL$ Status, and $MAF \times FRL$). The term for $SS \times FRL$ was significant for math scores in Grade 3 ($B_{SS \times FRL} = .0005$, $p < .01$), Grade 4 ($B_{SS \times FRL} = .0004$, $p < .01$), and Grade 5 ($B_{SS \times FRL} = .0008$, $p < .01$), indicating a stronger effect of SS on children who qualified for FRL than on children who did not qualify for FRL. The term for $MAF \times FRL$ was significant for math scores in Grade 3 ($B_{MAF \times FRL} = .0054$, $p < .01$), Grade 4 ($B_{MAF \times FRL} = .0068$, $p < .01$), and Grade 5 ($B_{MAF \times FRL} = .0064$, $p < .01$), indicating a stronger effect on FRL children than non-FRL children at all grades.

Analyses conducted separately for FRL-qualifying and nonqualifying children indicated that the effects of SS and MAF remained significant for both groups at every grade level (for SS: minimum $B_{SS} > .0061$, $p < .01$ for FRL children at each grade and minimum $B_{SS} > .0034$, $p < .001$ for non-FRL children at each grade; for MAF: minimum $B_{MAF} > .0111$, $p < .001$ for FRL children at each grade and $B_{MAF} > .0092$, $p < .001$ for non-FRL children at each grade). Appendix S4 shows that at all three grade levels, the impact of allocations for SS and MAF on math scores remained positive for both FRL-qualifying and FRL-nonqualifying

children, but the impact was stronger (i.e., the slope was steeper) for FRL-qualifying children for SS for Grades 4 and 5 and for MAF for Grades 3, 4, and 5. The figure shows that as funding allocations increased, the difference or "gap" between FRL-qualifying and FRL-nonqualifying children decreased.

Analyses of reading test scores revealed similar patterns. The interaction term for $SS \times FRL$ was significant for Grades 3 ($B_{SS \times FRL} = .0008$, $p < .01$), 4 ($B_{SS \times FRL} = .0011$, $p < .01$), and 5 ($B_{SS \times FRL} = .0011$, $p < .01$). The interaction term for $MAF \times FRL$ was significant for Grades 3 ($B_{MAF \times FRL} = .0028$, $p < .01$), 4 ($B_{MAF \times FRL} = .0069$, $p < .01$), and 5 ($B_{MAF \times FRL} = .0090$, $p < .01$). Again, follow-up analyses separately for FRL-qualifying and nonqualifying groups indicated significant effects of both SS and MAF for both groups at each grade level. The interaction effects indicated a difference in magnitude of effect favoring the FRL group in each case.

These findings indicate that SS had positive effects on reading and math test scores for children who qualified for FRL as well as for children who did not qualify, as we would expect given the program's focus on all children. Strikingly, MAF also had significant positive effects on children who qualified for FRL as well as for children who did not qualify and thus were unlikely to have received a direct MAF slot, indicating spillover effects of MAF on the rest of the population.

Analyses of special education placement revealed significant interaction effects between SS and FRL for Grades 3 ($B_{SS \times FRL} = .9922$, $p < .01$), 4 ($B_{SS \times FRL} = .9940$, $p < .01$), and 5 ($B_{SS \times FRL} = .9949$, $p < .01$). The interaction term for $MAF \times FRL$ was not significant for Grade 3 ($B_{MAF \times FRL} = .9939$, *ns*) but was significant for Grades 4 ($B_{MAF \times FRL} = .9747$, $p < .01$) and 5 ($B_{MAF \times FRL} = .9809$, $p < .01$). Follow-up analyses separately for FRL-qualifying and nonqualifying groups indicated significant effects of both SS and MAF for both groups at Grades 3, 4, and 5. The interaction effects indicated a difference in magnitude of effect favoring the FRL group.

Tests of Robustness Across Assumptions

The analyses reported thus far utilize the entire sample of children who matched a birth record to any educational outcome, no matter whether the county of birth matched the county of schooling. For these analyses, we assigned SS and MAF allocations to a child based on the county of the child's

birth. This analysis inevitably includes some children who moved across counties during the early childhood years and for whom the assigned SS and MAF allocations were inaccurate. We conducted three additional sets of analyses with different samples to test the robustness of findings across varying assumptions about where the child lived between birth and elementary school. To keep the number of analyses to a minimum, we used as dependent variables the child's mean reading and math scores in Grades 3, 4, and 5. Coefficients for all analyses are reported in Table 4.

The first robustness test utilized a limited sample of children for whom the birth county matched the child's school county in elementary school. We assumed that the child lived in the same county across the years from birth to school, although even this assigned allocation will have some errors for children who moved out and back into a county. We label this analysis "Same County" in Column 2 of Table 4. The second robustness test utilized the smaller sample of children who moved to a different county between birth and Grade 3 and for whom we assigned SS and MAF allocations based on the child's county of birth (labeled "Different County" in Column 3 of Table 4). The third

robustness test utilized the sample of children who moved to a different county between birth and Grade 3 and for whom we assigned SS and MAF allocations based on the Grade 3 school county (labeled "Different County" Column 4 of Table 4).

We found that both SS and MAF allocations were associated with significantly higher mean math scores and reading scores in Grades 3, 4, and 5, for 35 of the 36 analyses (Table 4). The R^2 coefficients indicate larger effect sizes for the same-county sample than the full sample, as one would expect if the numbers of mismatches in SS and MAF allocations are reduced. The R^2 coefficients indicate smaller effect sizes for the different-county sample analyses than the full sample or the same-county sample, as one would expect given the likely larger number of mismatches in assignment of the independent variables. Nonetheless, almost every significant finding remained significant in all robustness tests.

Discussion

The findings of this study indicate that state investments early in life in North Carolina's SS initiative

Table 4
Linear Regression Models on Math Scores and Reading Scores (Same County Versus Different County)

	(1) Full sample	(2) Same county	(3) Different county	(4) Different county
Math				
Grade 3				
Smart Start (\$00's)	.0051 (-.0013)***	.0049 (-.0015)***	.0058 (-.0012)***	.0013 (-.0005)**
More at Four (\$00's)	.0117 (-.0035)***	.0122 (-.0041)***	.0113 (-.0026)***	.0170 (-.0022)***
Grade 4				
Smart Start (\$00's)	.0050 (-.0012)***	.0053 (-.0015)***	.0045 (-.001)***	.0026 (-.0006)***
More at Four (\$00's)	.0170 (-.0033)***	.0196 (-.0039)***	.0119 (-.0029)***	.0179 (-.0025)***
Grade 5				
Smart Start (\$00's)	.0056 (-.0013)***	.0055 (-.0016)***	.0055 (-.0009)***	.0027 (-.0007)***
More at Four (\$00's)	.0204 (-.0036)***	.0250 (-.0043)***	.0109 (-.0032)***	.0183 (-.0023)***
Reading				
Grade 3				
Smart Start (\$00's)	.0069 (-.0011)***	.0071 (-.0013)***	.0065 (-.001)***	.0028 (-.0005)***
More at Four (\$00's)	.0144 (-.0026)***	.0142 (-.0032)***	.0159 (-.0026)***	.0178 (-.002)***
Grade 4				
Smart Start (\$00's)	.0062 (-.0011)***	.0066 (-.0013)***	.0051 (-.0009)***	.0034 (-.0005)***
More at Four (\$00's)	.0198 (-.0031)***	.0216 (-.0037)***	.0164 (-.0028)***	.0195 (-.0023)***
Grade 5				
Smart Start (\$00's)	.0077 (-.0012)***	.0084 (-.0014)***	-.0008 (.0173)***	-.0005 (.0186)***
More at Four (\$00's)	.0224 (-.0039)***	.0248 (-.0046)***	-.0029 (.0054)***	-.0018 (.0036)***

Note. Models include county and time fixed effects. Robust standard errors are in parentheses. All covariates presented in Table 1 are controlled in all models, but their coefficients are not listed. We use \$ from birth county in Column 3 and use \$ from school county in Column 4. ** $p < .05$. *** $p < .01$.

and MAF program are associated with higher math and reading standardized test scores, reductions in special education placement rates, and reductions in being grade retained in Grades 3, 4, and 5. The relation between financial investments and children's educational outcomes was linear, such that every hundred dollars of investment improved children's outcomes, at least throughout the range of funding in these programs. Because of fewer data points at the higher end, the confidence interval widens and conclusions must be more tentative. In the current range of funding, however, the findings indicate that every hundred dollars allocated per child would improve academic outcomes by about 0.001 *SD* for SS and 0.02 *SD* for MAF.

One could compare the effect sizes for the evaluation of SS and MAF with effect sizes reported for evaluations of other early childhood programs in other states. We report effect sizes in terms of the differences between children who live in a county in a year with the average dollar allocation for a program versus children living in a county in a year with no allocation. The effect sizes we report for SS are generally in the range of 0.07 *SD*, which are smaller than many effects reported for the Perry and Abecedarian programs; smaller than effects for community programs in Tulsa, New Jersey, and Tennessee; and similar to effect sizes after the 1st year of the Head Start Impact Study. However, unlike evaluations of other programs that include only those children who participate directly in a program, our design includes evaluation of impact on all children living in a county, even though many, if not most, children do not participate directly in an early childhood program and could benefit only through spillover or indirect effects. Given the crucial differences in design, the actual impact might not be any smaller than for those programs and might actually be more substantial.

We report the average effect size for MAF as about 0.20 *SD*, which is as large or larger than that for most pre-K programs. Numerous studies of other early childhood programs, particularly Head Start, have found that the impacts on children's cognitive and educational development fadeout by the end of elementary school. This study shows no such fadeout of effects of North Carolina's two flagship programs, SS and MAF, but instead shows impact that either holds steady or significantly increases through fifth grade. This pattern suggests that the impacts of early investments are not necessarily doomed to fadeout as cynics might believe but could actually lead to a trajectory of growing impacts across development, if implemented in the

right context. It becomes plausible to project continued benefit as children move into middle school. These findings support the likelihood of a greater benefit-cost ratio as the financial benefits continue across grade levels in reduced special education and grade retention costs. A full cost-benefit analysis is beyond the present scope but should be conducted for policy implications.

Although this study cannot identify precisely the reasons for the persistence and growth of impact of early childhood intervention, we point toward features of these programs and several additional findings in this study to speculate about the likely reasons. One reason commonly cited for fadeout is that treated preschool children subsequently enter elementary school environments populated by peers who had not experienced enriched early environments, leading to adverse peer influences and classrooms in which teachers spend inordinate time remediating peers instead of nurturing further growth in treated children. Indeed, these nontreated peers are sometimes used as the control group in evaluation studies. The North Carolina programs, in contrast, were intended to saturate communities. MAF did so by targeting every low-income child, whereas SS did so by addressing the community-wide quality of early childhood education and care. Both programs intended that kindergarten classrooms would be populated by entire cohorts of children ready to learn, so teachers could accelerate teaching and learning. Through processes of reciprocal influences in transactional development (Sameroff, 2009) and the technology of skill formation (Cunha & Heckman, 2007), perhaps the North Carolina programs changed the social ecology (Bronfenbrenner & Morris, 2006) of children's educational experience by having an impact on both the child and her or his peers. Indirect evidence to support this hypothesis is that each program was found to have positive impacts on children from both poverty and nonpoverty backgrounds, as if the entire cohort had benefitted. We note that we did not examine impact on children born out of state but entering NC schools to determine whether they benefitted from spillover. Future studies should examine the processes through which these early childhood programs exert impact over time and the conditions under which early impacts are sustained, perhaps through classroom observation and interviews with teachers in variously populated classrooms.

It is not clear why the measured effects of MAF appear larger than those for SS. Because MAF was implemented in the context of a state that had

already implemented SS and never without SS, we are cautious about comparing effect sizes across these programs; however, it is likely that the academic focus of MAF is responsible for its stronger academic impact. SS is a more holistic endeavor that might yield stronger impact on nonacademic outcomes such as health, well-being, and social competence. Future analyses should target a broader array of outcomes.

The impacts of SS and MAF were positive for all subgroups studied, and they were especially favorable for children receiving FRL. This greater impact on disadvantaged children is most likely due to the fact that they were targeted for early childhood intervention. SS allocated extra child-care subsidies for low-income children, and MAF required low family income for support. It could also be explained by limited range in the measurement of special education and grade retention outcomes in this study that limited the upper bound of measurement of impact on some children. Yet another possibility is that low-income children are more responsive to the same intervention than are middle-income children (Belsky & Pluess, 2009). That is, it could be that the same conditions that lead some children to be highly negatively affected by adverse conditions could lead them to be highly positively affected by enriched conditions. Future studies should extend the measurement of outcomes to examine more closely the impacts on all children.

Even though the MAF Program (and, to some extent, SS) targeted low-income children, we found evidence of likely spillover of impacts on children who do not qualify for FRL. It is plausible that positive impacts on low-income children led school teachers to focus more on teaching all students at grade level rather than to attend to behavior problems or remediation, thus enhancing academic learning for all students. It is also plausible that middle-income students benefitted from exposure to positively performing low-income children. Yet another possibility is that because some nonqualifying third graders may have been poor at a younger age and may have qualified for MAF, the apparent spillover effect might instead be a direct effect for this subgroup of students who changed poverty status.

Although the findings of this study are robust across measures, age levels, and different assumptions about samples, confidence in the causal role of these two early childhood programs is tempered by an important feature of the evaluation design. The design assumes no unmeasured third variable

that is responsible for both a county's funding level in a particular year and subsequent outcomes for the cohort of children of that age in that county in that year. We controlled for secular changes in outcomes through time fixed analyses, stable differences in counties through county fixed effects analyses, and time-varying county characteristics and child-level characteristics through covariance analyses. But it remains plausible that an unmeasured variable is correlated with financial allocations to a county in a particular year. Informal interviews with state leaders and local stakeholders suggested that a charismatic or persuasive community leader such as an elected official might have tried to secure higher than expected state funding in a particular year. Although it is plausible that the same official might have independently inspired other local improvements in early child care, it is unlikely that this phenomenon occurred systematically across counties and years. It is even less likely that these leaders had an impact only on a particular age cohort within a community, and so her or his cross-cohort impact would be controlled by county fixed effects.

These findings provide support for both developmental science rationales for early childhood programs. That is, the SS initiative was predicated on the need for improvement in early child-care quality, especially for low-income children. We have found that this initiative leads to improvements in children's educational outcomes through age 11. Likewise, the MAF program was predicated on the need to provide high-quality preschool for low-income children in the year prior to kindergarten, and we have found that this program also leads to improvements through the end of elementary school. Larger impacts for MAF suggest that an approach that focuses specifically on academic skills in the year just prior to school entry will have greater impact on academic outcomes. SS broader approach might have larger impacts on other domains (e.g., health, social competence) that were not measured here. We also found support for the developmental perspective that academic performance for disadvantaged children can be enhanced and sustained if both early intervention and a later environment of positive peer performance are implemented.

The policy implications of these findings are several. First, each of these two programs would seem to merit dissemination and continued support by government. Second, the sustained effects suggest that it may be more beneficial to saturate a community with an early childhood program than to

distribute limited resources across many communities at a level that is below a threshold of enduring impact. Future research should examine whether levels of penetration across a community affect fadeout of impact over time. Third, the findings beg for funding of comprehensive benefit–cost analyses of early childhood programs that include a broad array of educational outcomes, given the current findings of impacts on test scores, special education placements, and grade retention, all of which might have financial value; this benefit–cost analysis should incorporate the entire population in order to capture spillover effects on peers.

In sum, the findings of this analysis of North Carolina's two early childhood programs indicate that full implementation leads to significant impacts that persist through the end of elementary school.

References

- Andrews, R. J., Jargowsky, P., & Kuhne, K. (2012). *The effects of Texas's targeted pre-Kindergarten program on academic performance*. NBER Working Paper No. 18598.
- Bailey, D., Duncan, G., Odgers, C., & Yu, W. (2015). *Persistence and fadeout in the impacts of child and adolescent interventions*. Working paper, Irvine Network on Individuals in Development, University of California, Irvine, CA.
- Barnett, W. S., Carolan, M. E., Squires, J. H., & Clarke Brown, K. (2013). *The state of preschool 2013: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- Barnett, W. S., Carolan, M. E., Squires, J. H., Clarke Brown, K., & Horowitz, M. (2015). *The state of preschool 2014: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- Barnett, W. S., Jung, K., Youn, M.-J., & Frede, E. C. (2013). *Abbott Preschool Program Longitudinal Effects Study: Fifth grade follow-up*. New Brunswick, NJ: National Institute for Early Education Research.
- Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, *135*, 885–908. doi:10.1037/a0017376
- Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics*, *130*, e1136–e1145. doi:10.1542/peds.2012-0243
- Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Theoretical models of human development. Vol. 1: The handbook of child psychology* (5th ed., pp. 793–828). New York, NY: Wiley.
- 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, CA: University of North Carolina, FPG Child Development Institute.
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, *112*, 579–620.
- Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, *65*, 684–698. doi:10.1111/j.1467-8624.1994.tb00777.x
- Cunha, F., & Heckman, J. J. (2007). The technology of skill formation. *American Economic Review*, *97*(2), 31–47.
- Curenton, S. M., Dong, N., & Shen, X. (2015). Does aggregate school-wide achievement mediate fifth grade outcomes for former early childhood education participants? *Developmental Psychology*, *51*, 921–934. doi:10.1037/a0039295
- Davoudzadeh, P., McTernan, M. L., & Grimm, K. J. (2015). Early school readiness predictors of grade retention from kindergarten through eighth grade: A multi-level discrete-time survival analysis approach. *Early Childhood Research Quarterly*, *32*, 183–192.
- Dodge, K. A., & Haskins, R. (2015). Children and government. In R. M. Lerner, M. H. Bornstein, & T. Leventhal (Eds.), *Handbook of child psychology and developmental science, volume four, ecological settings and processes* (7th ed., pp. 654–703). New York, NY: Wiley.
- Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development*, *65*, 649–665. doi:10.1111/j.1467-8624.1994.tb00774.x
- 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), Article 46.
- Gormley, W. T. J., & Gaye, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program. *Journal of Human Resources*, *40*, 533–558. doi:10.3368/jhr.XL.3.533
- Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*, 527–544. doi:10.3386/w8502
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.
- Haskins, R., & Rouse, C. (2005). Closing achievement gaps. *The Future of Children*. Retrieved from http://www.futureofchildren.org/publications/docs/15_01_PolicyBrief.pdf.
- Justice, L. M., Petscher, Y., Schatschneider, C., & Mashburn, A. (2011). Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills? *Child Development*, *82*, 1768–1777. doi:10.1111/j.1467-8624.2011.01665.x
- Ladd, H. F., Muschkin, C., & Dodge, K. A. (2014). From birth to school: Early childhood initiatives and third-

- grade outcomes in North Carolina. *Journal of Policy Analysis and Management*, 33, 162–187. doi:10.1002/pam.21734
- Lipsey, M. W., Farran, D. C., & Hofer, K. G. (2015). *A randomized control trial of the effects of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade (Research Report)*. Nashville, TN: Vanderbilt University, Peabody Research Institute.
- McLanahan, S., & Jacobsen, W. C. (2015). Diverging destinies revisited. In P. R. Amato, A. Booth, S. M. McHale, & J. Van Hook (Eds.), *Families in an era of increasing inequality: Diverging destinies* (pp. 3–23). New York, NY: Springer.
- Mrazek, P. J., & Haggerty, R. J. (Eds.). (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington, DC: National Academy Press.
- Muschkin, C. G., Ladd, H. F., & Dodge, K. A. (2015). Impact of North Carolina's early childhood initiatives on special education placements in third grade. *Education Evaluation and Policy Analysis*, 37, 478–500.
- N.C. Office of State Budget and Management. (2014). *County/state population projections*. Retrieved from http://www.osbm.state.nc.us/ncosbm/facts_and_figures/socioeconomic_data/population_estimates/county_projections.shtm
- Peisner-Feinberg, E. S.; the More at Four Evaluation Team. (2008). *Evaluation of the North Carolina more at four pre-kindergarten program, year 6*. Chapel Hill, NC: The University of North Carolina, Frank Porter Graham Child Development Institute.
- Sameroff, A. J. (Ed.). (2009). *The transactional model of development: How children and contexts shape each other* (pp. 117–135). Washington, DC: American Psychological Association.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005a). *Lifetime effects: The high/scope perry preschool study through age 40* (Monographs of the High/Scope Educational Research Foundation, 14). Ypsilanti, MI: High/Scope Educational Research Foundation.
- Schweinhart, L., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005b). *Lifetime effects: The high/scope perry preschool study through age 40*. Ypsilanti, MI: High Scope Press.
- Schweinart, L. J., Xiang, Z., Daniel-Echols, M., Browning, K., & Wakabayashi, T. (2012). *Michigan Great Start Readiness Program Evaluation 2012: High school graduation and grade retention findings*. Prepared for the Michigan Department of Education. Retrieved from http://www.highscope.org/file/Research/state_preschool/MGSRP%20Report%202012.pdf
- Shonkoff, J. P. (2014). A healthy start before and after birth: Applying the biology of adversity to build the capabilities of caregivers. In K. McCartney, H. Yoshikawa, & L. B. Forcier (Eds.), *Improving the odds for America's children* (pp. 28–39). Cambridge, MA: Harvard Education Press.
- Temple, J. A., & Reynolds, A. J. (2007). Benefits and costs of investments in preschool education: Evidence from the Child-Parent Centers and related programs. *Economics of Education Review*, 26(1), 126–144. doi:<http://dx.doi.org/10.1016/j.econedurev.2005.11.004>
- U.S. Department of Health and Human Services, Administration for Children and Families. (2010, January). *Head start impact study*. Final Report. Washington, DC.
- Yeung, W. J., Linver, M. R., & Brooks-Gunn, J. (2002). How money matters for young children's development: Parental investment and family processes. *Child Development*, 73, 1861–1879. doi:10.1111/1467-8624.t01-1-00511
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., . . . Zaslow, M. J. (2013). *Investing in our future: The evidence base on pre-school education*. Retrieved from www.srcd.org/policy-media/policy-updates/meetings-briefings/investing-our-future-evidence-base-preschool

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Descriptive Characteristics of Research Participants.

Appendix S2. Linear Regression Model Results for Math Scores.

Appendix S3. Linear Regression Models for Reading Scores.

Appendix S4. Interaction Effects Between Free or Reduced Lunch Status and Smart Start (and More at Four) Funding Allocations on Math Test Scores in Grades 3, 4, and 5.

Appendix S5. Group Means and Tests of Differences between Children Whose Birth County and Grade 3 School County Were the Same and Children Whose Birth County and Grade 3 School County Differed.