

Head Start's Comparative Advantage: Myth or Reality?

William T. Gormley Jr., Deborah Phillips, Shirley Adelstein, and Catherine Shaw

In recent years, Head Start's prominent role in preparing low-income 4-year-olds for school has been affected by rapid growth in state-funded pre-K programs, some of which are based in public schools. This has led to questions about the comparative advantages of these two approaches to early education. An analysis of data from Tulsa, Oklahoma, indicates that the school-based pre-K program is more effective in improving early literacy outcomes, while Head Start is more effective in improving health outcomes. The two programs are comparable with regard to early math learning. Social-emotional effects are more subtle, but the school-based pre-K program has demonstrable positive effects, while the Head Start program does not.

For many years, the Head Start program enjoyed a virtual monopoly over government-funded early childhood education services for disadvantaged 4-year-old children. This began to change in the early 1990s, as new federal legislation channeled funds to disadvantaged children attending day care centers and family day care homes. Since the late 1990s, the landscape of early education programs has become even more crowded and more complex. Not only are efforts to improve the capacity of childcare programs to prepare young children for school on the rise, but state-funded pre-K programs, which now serve more 4-year-olds than Head Start (Barnett, Epstein, Friedman, Boyd, & Hustedt, 2008), are assuming a very prominent place on this landscape. Moreover, as with Head Start, most of the 4-year-olds served by state pre-K programs are disadvantaged.

Are disadvantaged 4-year-olds better served by Head Start or by state-funded pre-K programs? What are the comparative advantages of these two approaches—one comprehensive in its goals, the other more clearly focused on early learning—for preparing young children for school? These questions are of considerable interest to both public officials and parents. From the public officials' perspective, Head Start is superior to state-funded pre-K only if its multiple benefits are high enough to justify its higher costs. Parents necessarily seek early education services for their children with a broad array of needs in mind, ranging from cultural compatibility to coverage of work hours. Nevertheless, for those whose primary goal is to ensure that their children get off to a good start in kindergarten, Head Start is the better bet only if

it produces better educational outcomes or some other combination of positive outcomes for their children.

In this paper, we consider evidence from Tulsa, Oklahoma, which boasts both a high-quality Head Start program and a high-quality state-funded pre-K program. In addition to reanalyzing previously reported evidence on the two programs' cognitive and social-emotional effects, we report fresh evidence on the two programs' health effects. Health services are among the most important "noneducational" services that the Head Start program provides. If Head Start programs produce favorable health outcomes, while pre-K programs do not, this might be one reason to continue to support a dual system of delivering educational services to disadvantaged children. If not, we may wish to consider other alternatives, such as giving the states greater discretion to consolidate or integrate Head Start and pre-K programs.

The Goals of Early Childhood Programs

When the Head Start program was created in 1965, it was designed as a "comprehensive" program to help young children from poor families. Program objectives encompassed the fields of health, social services, and education (Zigler & Muenchow, 1992, p. 18). The Head Start program's seven objectives included: (i) improving the child's physical health and physical abilities; (ii) helping the emotional and social development of the child by encouraging self-confidence, spontaneity, curiosity, and self-discipline; (iii) improving the child's mental processes and skills, with particular attention to conceptual and verbal skills; (iv) establishing patterns and expectations of success for the child that will create a climate of confidence for future learning efforts; (v) increasing the child's capacity to relate positively to family members and others while at the same time strengthening the family's ability to relate positively to the child and his problems; (vi) developing in the child and his family a responsible attitude toward society, and encouraging society to work with the poor in solving their problems; and (vii) increasing the sense of dignity and self-worth within the child and his family (Richmond, Stipek, & Zigler, 1979, p. 37).

Since 1965, the Head Start program's mission has changed in some respects, with an even greater emphasis on early childhood education. This is evident in recent congressional mandates to improve the educational credentials of Head Start teachers. Most notably, Congress required that by September 30, 2003, at least half of all Head Start teachers in center-based programs must have an AA, BA, or advanced degree in early childhood education or a degree in a related field, with preschool teaching experience. Later, Congress required that 50 percent of teachers have a BA degree in early childhood education or a related field, and that all new hires have an AA degree or be on track to receive an AA degree within 3 years.

The best way to describe Head Start's current status is that it is a school readiness program with a relatively broad conception of what constitutes school readiness. According to the Administration for Children and Families, Head Start is "a national program that promotes school readiness by enhancing the social and cognitive development of children through the provision of educational, health, nutritional, social and other services to enrolled children and families" (U.S. Department of

Health and Human Services, Administration for Children and Families, 2008). In contrast, state-funded pre-K programs serve a more purely educational function. In Oklahoma, for example, the 1998 legal mandate establishing the state's universal pre-K program says nothing about health services, family support services, or parental involvement. Some public schools in Oklahoma, as elsewhere, do provide health services, typically to disadvantaged children, but this is not required by law.

Literature Review

Early intervention (EI) programs for children in poverty seek to compensate for suboptimal home environments, especially inadequate resources to prepare young children for the cognitive and social demands of elementary school. A longstanding and extensive body of evidence suggests that this goal can be achieved (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Reynolds, Temple, Robertson, & Mann, 2001; Schweinhart et al., 2005). However, many of these carefully constructed, high-quality, and costly programs do not reflect the early childhood options—including state pre-K programs—that are typically available to low-income families today, nor do the comparison groups used in these studies reflect the dramatically different counterfactuals that now characterize the peers of children who attend EI programs.

Fortunately, a new literature on both school-based pre-K and Head Start is now emerging that can guide contemporary hypotheses about how these programs are likely to affect children's health and social and cognitive development, for both poor and nonpoor children (Gormley, 2007). In general, this literature is most informative with regard to cognitive effects, less so about social-emotional effects, and still less so with regard to health effects.

Studies using a national data set (Early Childhood Longitudinal Study–Kindergarten; ECLS–K) have found a modest positive association between participation in pre-K, as defined by parents, or, more broadly, center-based care, and cognitive test scores in kindergarten (Loeb, Bridges, Bassok, Fuller, & Rumberger, 2006; Magnuson, Ruhm, & Waldfogel, 2007). State-specific studies have also documented positive effects on language skills and math skills (Henry et al., 2004; Wong, Cook, Barnett, & Jung, 2008; Xiang & Schweinhart, 2002). City-specific studies have found substantial short-term improvements in cognitive outcomes in Tulsa, Oklahoma (Gormley, Gayer, Phillips, & Dawson, 2005; Gormley, Phillips, & Gayer, 2008) and in Chicago, Illinois (Reynolds & Temple, 1995).

The Head Start Impact Study, a randomized experiment funded by the federal government, found evidence of positive, but modest, short-term cognitive impacts for 3- and 4-year-olds, though not in every area (Puma, Bell, Cook, Heid, & Lopez, 2005). For 3-year-olds, Head Start produced statistically significant improvements for two of two pre-reading tests, one of two pre-writing tests, and two of two vocabulary tests. For 4-year-olds, Head Start yielded statistically significant gains for two of two pre-reading tests, one of two pre-writing tests, and neither of the two vocabulary tests. Head Start had no impact on oral comprehension, phonological awareness, or early math skills for 3- or 4-year-olds. Ludwig and Phillips (2008) obtained somewhat

stronger effects for these outcomes in their reanalysis of the Impact Study data using an approach that takes into account the fact that some experimental children did not actually enroll in Head Start and some control children did receive Head Start services. However, the latest iteration of the Head Start Impact Study reveals diminished impacts on cognition by the end of first grade (Puma, Bell, Cook, & Heid, 2010).

In our research in Tulsa, we have found even stronger impacts for the Community Action Project (CAP) of Tulsa County Head Start program on children's early literacy and math skills at kindergarten entry (Gormley et al., 2008). Importantly, a greater share of time is spent on instructional activities in the Tulsa Head Start program than is the case in other Head Start programs (Phillips, Gormley, & Lowenstein, 2009). Head Start studies using sibling comparisons have found evidence of long-term positive impacts on academic achievement, though not for every racial and ethnic group (Currie & Thomas, 1995, 1999; Garces, Thomas, & Currie, 2002).

Both Head Start and school-based pre-K programs have also been documented to have impacts on early social development, although these impacts are not always positive, are not always consistent, and are seldom strong. Studies using a national data set (ECLS-K) have found a relationship between pre-K participation and more externalizing behavior and less self-control in kindergarten (Loeb et al., 2006; Magnuson et al., 2007), suggesting negative social-emotional effects from pre-K participation. The one exception to this pattern occurred when the children's pre-K and kindergarten classrooms were co-located in the same school. State-specific and city-specific studies of the effects of pre-K have typically not employed social-emotional measures, largely because of measurement challenges.

The Head Start Impact Study shows a reduction in overall problem behaviors and hyperactivity for 3-year-olds (U.S. Department of Health and Human Services, 2005), but no impacts were found on parent-reported social skills, positive approaches to learning, or parent-reported social competence. No impacts were found for the 4-year-olds in the study. The latest iteration of the Head Start Impact Study shows diminished social-emotional effects at the end of first grade (Puma et al., 2010). Nonexperimental national studies found negative effects of Head Start participation on social-emotional behavior among kindergarteners (Loeb et al., 2006; Magnuson et al., 2007).

In our own work in Tulsa (Gormley, Phillips, Newmark, Welti, & Adelstein, 2011; Lowenstein, 2009), participation in the Tulsa Public Schools (TPS) pre-K program was associated with lower teacher ratings on timidity and higher ratings on attentiveness. Participation in the Head Start program had no discernible effects. We found similar, but not identical, results when we restricted the TPS and Head Start samples to free lunch-eligible children.

Thanks to the Head Start Impact Study, we have good information on the short-term effects of Head Start on children's health. For overall health status, as perceived by parents, Head Start had a positive, statistically significant effect for 3-year-olds (the effect size was 0.12), but no statistically significant effect for 4-year-olds (Puma et al., 2005). Effects on dental care were more striking, with effect sizes of 0.34 for 3-year-olds and 0.32 for 4-year-olds. To put it more concretely, 73 percent of the treatment group had dental care, as opposed to 57 percent of the control group,

among 4-year-olds; 69 percent of the treatment group had dental care, as opposed to 52 percent of the control group, among 3-year-olds (Puma et al., 2005, p. xv). As with the other indicators, the latest iteration of the Head Start Impact Study shows diminished health impacts at the end of first grade (Puma et al., 2010). Ludwig and Miller (2007), focusing on some of the earliest Head Start participants, found that a 50–100 percent increase in Head Start funding is associated with a decline in mortality from causes of death that could be affected by the program of 33–50 percent of the control mean.

Evidence on the effects of school-based health care programs is more limited, but growing. A recent study in Newark, New Jersey, found no general impact of school-based clinics on emergency room use or hospitalization but a modest positive impact on the likelihood of using health care, including medical, dental, and mental health care (Silberberg & Cantor, 2008). In contrast, other studies have found that school-based clinics do reduce hospitalizations and visits to the emergency room (Kaplan et al., 1999; Santelli, Kouzis, & Newcomer, 1996; Young, D'angelo, & Davis, 2001). A few studies have focused on asthma-related outcomes, as this is the most common chronic disease among children, and these studies have similarly found that school-based clinics are associated with lower hospitalization rates and outpatient visits (Lurie, Bauer, & Brady, 2001; Webber et al., 2003). Unfortunately, few studies have examined the effects of school-based clinics on overall health status. One study of health centers in urban secondary schools found no significant effects on overall health status (Kisker & Brown, 1996), but little is known about health status effects for younger children. A common challenge in all of these studies is that because school-based health centers are typically established at schools with the neediest students, it is often hard to find appropriate comparison schools with similar demographics, but no health center. At this point, it is best to characterize the evidence on school-based health care for younger children as inconclusive, especially with regard to overall health status.

Tulsa's Early Childhood Programs

The Tulsa Public Schools pre-K program satisfies all state requirements specified in Oklahoma's 1998 universal pre-K law. All lead teachers must have a BA degree and are early childhood certified. They are paid a regular public school wage, like teachers in the higher grades. Thanks to assistant teachers, their classrooms meet a 10/1 child/staff ratio requirement (or lower). At the time of our study, TPS received approximately \$3,700 per student for every student enrolled full time in pre-K (and approximately \$2,000 per student for every student enrolled half time in pre-K). At the time of our study, 61 percent of TPS pre-K programs were full time.

The second largest early childhood program in Tulsa County is the CAP of Tulsa County Head Start program. Under the 1998 law, Head Start programs may collaborate with local school districts if both partners agree; TPS and CAP of Tulsa County have collaborated for several years. Under this arrangement, CAP of Tulsa County Head Start receives 90 percent of the funding per student that would otherwise flow to TPS if there were no collaboration (TPS gets the rest). Like TPS, all lead teachers

at CAP Head Start have a BA degree and are early childhood certified. To compete for the most capable teachers, CAP Head Start's leaders have chosen to pay teachers the same wages and benefits that they would receive from TPS (or slightly higher). As with TPS, a 10/1 child/staff ratio is maintained. All of CAP Head Start's 4-year-old programs are full time.

It should be noted that, based on the usual criteria (education, wages, and child/staff ratios), both the TPS pre-K program and the CAP of Tulsa County Head Start program are better than their respective averages (all school-based pre-K programs; all Head Start programs). Both programs are also better than average if one examines more sophisticated measures of program quality, based on in-depth classroom observations (Phillips et al., 2009).

Data

The data available to us, never before discussed in one article, come from four sources: cognitive testing conducted in Tulsa, OK in August 2006, a survey of the parents who accompanied their child to that testing, social-emotional assessments of the child's teacher in early October 2006, and administrative data from Tulsa Public Schools and the CAP of Tulsa County Head Start program. The cognitive testing and the parent survey took place just prior to the commencement of classes, while the social-emotional assessments took place 40 days into the semester. For reasons discussed below, we assessed children beginning the Tulsa pre-K program, children beginning the CAP of Tulsa County Head Start program, and children beginning kindergarten in the Tulsa Public Schools.

We selected as our measures of cognitive development three subtests of the Woodcock-Johnson Achievement Test: the Letter-Word I.D. Test (a measure of pre-reading skills), the Spelling Test (a measure of pre-writing skills), and the Applied Problems Test (a measure of pre-math skills). These subtests were chosen because they are particularly appropriate for relatively young children. The tests were administered by the child's new teacher, in a one-on-one meeting that took place just before classes commenced. Teachers were trained to administer these three subtests by Barbara Wendling of Dallas, Texas, an expert on the Woodcock-Johnson Test. While the child was tested, the parent completed the questionnaire survey, which asked questions about the child's background and the family's background. A majority of Hispanic children were tested in both English and Spanish (Gormley, 2008), but only the English results are reported here.

To measure children's social-emotional development, we relied on an assessment tool developed by researchers at the University of Pennsylvania, known as the Adjustment Scales for Preschool Intervention (ASPI). We also included four questions from a self-regulation scale that seeks to capture the child's attentiveness, which has been shown to be a good predictor of later achievement (Duncan et al., 2007). Unlike cognitive development tests, which can be administered by teachers to new students who have had no prior contact with the testing teacher, social-emotional assessments require substantial knowledge of the student's habits and behavior. Following standard practice, we asked teachers to assess their students'

social-emotional development approximately 40 days into the school year. We later converted the teachers' assessments on the ASPI scale into five factors (or phenotypes), using common factor analysis and orthogonal equamax rotation for the loadings. The five factors were standardized to have a mean of 50 and a standard deviation of 10 (Gormley et al., 2011).

Finally, to measure health effects, we relied on three health questions from the parent survey. The first question asked the parent to rate the child's health using a standard Likert scale: excellent, very good, good, fair, or poor. The second question asked the parent to specify whether the child's last visit to a clinic, health center, hospital, doctor's office, or other place for routine health care took place within the last 6 months, between 6 months and 1 year earlier, between 1 and 2 years earlier, more than 2 years earlier, or never. The third question asked the parent to specify whether the child's last visit to a dentist or dental hygienist took place within the last 6 months, between 6 months and 1 year earlier, between 1 and 2 years earlier, more than 2 years earlier, or never.

For the cognitive testing, approximately 78 percent of all pre-K entrants, 69 percent of all Head Start entrants, and 73 percent of all kindergarten entrants were tested. This yielded sample sizes of 1,546, 510, and 3,000, respectively. The tested students closely resembled the universe of students, except for the kindergarten cohort, where some small differences between the two groups were found.¹ We received parent surveys from 86 percent of tested students.

For the social-emotional assessments, we received completed forms for 77 percent of the kindergarten students. This yielded a sample of 3,166 kindergarteners. The assessed students closely resembled the universe of students.² We received parent surveys from 73 percent of assessed kindergarten students. As explained below, only the kindergarten children were relevant for the social-emotional assessments, though the other children were also assessed.

For the health assessments, our outcome measures are derived directly from the parent surveys. We received parent surveys with valid responses for at least one health-related variable from 62 percent of the kindergarten students. This yielded a sample of 2,548 kindergarteners. The assessed students closely resembled the universe of students.³ As with the social-emotional assessments, only the kindergarten students were relevant for the health assessments because of our estimating technique described below.

For all analyses, our analytical sample was limited to students who would be age-appropriate for their class if the birthday cut-off were perfectly enforced.⁴ For the cognitive testing, this includes approximately 77 percent of all pre-K entrants, 69 percent of all Head Start entrants, and 66 percent of all kindergarten entrants. This yielded sample sizes of 1,515, 505, and 2,701, respectively. For the social-emotional assessments, this includes approximately 69 percent of all K entrants, yielding a sample size of 2,829. For the health assessments, this includes approximately 56 percent of all K entrants, yielding a sample size of 2,317.

Some of our variables were almost never missing (gender, age, race/ethnicity, free lunch eligibility), but some variables from our parent survey (Internet access, marital status, biological father lives at home, parent's place of birth, number of

books at home) were missing for about 40 percent of our sample, and two variables from our parent survey (mother's education, childcare history) were missing even more frequently. For all analyses, we handled missing data through the multiple imputation method (Rubin, 1987), which involves creating multiple data sets in which missing values are imputed based on observed data. These complete data sets are then analyzed separately, and the results are combined to produce final parameter estimates and standard errors. By combining results across multiple imputations, this approach incorporates the uncertainty associated with the imputation of missing data. Although somewhat cumbersome, multiple imputation has been shown to perform better than other common methods of addressing missing data (Croy & Novins, 2005; Rubin, 1996; Sinharay, Stern, & Russell, 2001).

Methodology

One of the greatest challenges that researchers face when evaluating program impacts is selection bias. This problem arises because individuals who receive treatment often differ systematically from individuals in the nonexperimental comparison group, biasing the estimated treatment effect. In a series of articles on the Tulsa Public Schools pre-K program, we have used a regression-discontinuity design to estimate the impacts of the Tulsa pre-K program on children's cognitive development (Gormley & Gayer, 2005; Gormley et al., 2005, 2008). We have used a similar strategy to estimate the impacts of the Tulsa Head Start program on children's cognitive development (Gormley et al., 2008).

A regression-discontinuity design, applied to early childhood education data, takes advantage of a strict September 1 birthday requirement for enrolling children in the state of Oklahoma's pre-K program for 4-year-olds. A child born on September 2, 2001, or later, was ineligible for enrollment in the 2005–06 school year; a child born on September 1, 2001, or earlier, was eligible to enroll. By comparing children who just completed the pre-K program with children who just began the pre-K program, and by controlling for their precise date of birth (and other variables), we can obtain a relatively unbiased estimate of program impact (Gormley & Gayer, 2005). That is because both groups of children have parents who affirmatively chose to enroll their child in the same pre-K program. As a test of whether we are comparing similar children, we can compare the observable demographic characteristics of both sets of children. When we do so, we see striking similarities between the two groups of children, which gives us confidence that we have chosen a suitable comparison group.

When thinking about other dependent variables, however, we have not always concluded that a regression-discontinuity design is the most appropriate strategy. As noted earlier, we used the ASPI instrument to measure social–emotional development, which required us to wait 40 days into the school year for the teacher's assessment. Unlike the cognitive testing, this is not a pure pretest. Were we to apply the regression-discontinuity design to social–emotional development, we would in effect be comparing students who completed one year of pre-K and 40 days of kindergarten with students who completed 40 days of pre-K. A better strategy, we

thought, was to compare students who completed pre-K and 40 days of kindergarten with students who completed 40 days of kindergarten and no pre-K. An additional problem with applying the regression-discontinuity design to social-emotional development is that kindergarten teachers may employ higher standards for maturity than pre-K teachers, because they are accustomed to dealing with kindergarten students. If so, students in the treatment group would be held to a higher standard than students in the comparison group. For both of these reasons, we decided to use propensity score matching when assessing the effects of program participation on children's socio-emotional development (Gormley et al., 2011). With propensity score matching, we can compare kindergarten children who completed pre-K with kindergarten children who did not attend pre-K. Similarly, we can compare kindergarten children who completed Head Start with kindergarten children who did not attend Head Start. As an additional precaution, we used a teacher fixed effects model to guard against possible teacher bias. We also included demographic covariates in this model. Using OLS regression with teacher fixed effects allows us to account for the possibility that teachers differentially applied the ASPI instrument, or that some teachers improved the social-emotional development of their children more than others during the first 40 days of the school year.

To assess the effects of the Tulsa pre-K program (and the Tulsa Head Start program) on children's health, we included three health-related questions in a parent survey administered while the student was being tested for cognitive development. As noted earlier, the first question asked the parent to assess the child's overall health, the second question asked the parent to specify how recently the child had visited a physician, and the third question asked the parent to specify how recently the child had visited a dentist. As we considered the regression-discontinuity design in this context, it occurred to us that the likelihood of a child's visiting a physician to deal with a health problem might be influenced by age-specific immunization requirements.⁵ If the program promotes or facilitates doctor visits, then it becomes difficult to disentangle doctor visits due to immunization schedules from doctor visits due to greater contagion (negative effects) or greater parental awareness (positive effects). If we compare children who are in the same age cohort, these immunization schedule effects are less likely. We also concluded that while parents may take their child's educational ability into account when trying to decide whether to enroll the child in pre-K or Head Start, they are less likely to take their child's health status into account when trying to decide whether to enroll the child in pre-K or Head Start. If so, the selection bias problems that the regression-discontinuity design seeks to address may be less worrisome in this context. For all these reasons, we decided to use propensity score matching and a linear probability model when assessing program impacts on children's health.⁶ We could not control for parental bias in the same way that we controlled for teacher bias when assessing social-emotional outcomes, because each parent filled out only one survey.⁷

Propensity score matching enables us to estimate a causal effect by identifying a control group that closely resembles the treatment group with regard to observable characteristics. Members of the treatment and control groups are matched based on similarity of propensity scores, which capture the likelihood of being in the treat-

ment group conditional on a wide range of measured characteristics. Assuming that selection into treatment is a function of observed characteristics, this method can effectively reduce selection bias by comparing the outcomes of treatment and control groups that are as similar as possible (Dehejia & Wahba, 1999, 2002; Rosenbaum & Rubin, 1983, 1985).

Our method involved four steps. First, we used a wide variety of covariates to estimate a logistic regression predicting the probability of having attended TPS pre-K the previous year for the sample of TPS kindergarten students who either attended TPS pre-K the previous year or attended neither TPS pre-K nor Head Start.⁸ Second, the estimated regression results were used to generate a propensity score—the predicted probability of having attended TPS pre-K the previous year—for each observation in the sample. Third, we employed the PSMATCH2 program in Stata to match treatment and control observations based on their propensity scores (Leuven & Sianesi, 2003). Finally, we used a linear probability model to estimate program effects for the samples produced by propensity score matching. The covariates in this model included race, gender, free lunch eligibility, mother's education, whether the child lives with his or her biological father, and whether the child has Internet access at home. To implement multiple imputation for missing data, we used the ice program in Stata to create five imputed data sets (Royston, 2004, 2005a, 2005b). We applied our propensity score matching technique to each of these imputed data sets and used the micombine command to produce our final regression results. This procedure involves estimating separate regressions for each of the five matched samples. Final parameter estimates reflect averages across these regression analyses, and standard errors are calculated following the rules developed by Rubin (1987). An identical approach was used to estimate Head Start treatment effects, but in this case, the sample included TPS kindergarten students who either attended Head Start the previous year or attended neither TPS pre-K nor Head Start, and the propensity scores reflected the predicted probability of having attended Head Start the previous year. When pairing treatment and control groups, there are several possible matching techniques. For our purposes, we employed nearest-neighbor one-to-one matching with replacement. We also imposed a caliper of 0.001 for the estimation of TPS pre-K treatment effects and 0.005 for the estimation of Head Start treatment effects. This matching procedure begins by randomly sorting the treatment and control observations. The first treatment observation is then matched with the control observation with the closest propensity score, and the matching process continues in order for the remaining treatment observations.

Matching with replacement allows a control observation to be matched with multiple treatment observations if he or she has the closest propensity score. This reduces bias by improving the quality of the match, and control observations that are not matched are dropped from the analysis. Imposing a caliper similarly improves the quality of the match by limiting the allowable difference in propensity scores between matched pairs. However, in the presence of a caliper, treated observations that are not matched are dropped from the analysis, resulting in the exclusion of a larger number of observations from the estimation of treatment effects. We imposed different calipers in our TPS pre-K (0.001) and Head Start (0.005) analyses because it

was more difficult to find suitable matches for the Head Start sample. In practice, this means that we have somewhat greater confidence in our TPS matching than in our Head Start matching (but considerable confidence in both).

Following this methodology for the social-emotional phenotype analyses, across five imputations 1,128–1,167 TPS pre-K treatment observations were matched with 566–587 unique comparison observations, and 326–342 CAP Head Start treatment observations were matched with 215–222 unique comparison observations. For our attentiveness index analyses, 1,105–1,141 TPS pre-K alumni were matched with 554–576 unique comparison observations, and 320–336 CAP Head Start alumni were matched with 210–219 unique comparison observations. For our analysis of health outcomes, matched sample sizes across the five imputations varied slightly among the outcome variables, because in some cases, a parent responded to some but not all of the three health-related survey items. For our health status analyses, 895–929 TPS pre-K treatment observations were matched with 471–489 unique comparison observations, and 236–246 CAP Head Start treatment observations were matched with 159–168 unique comparison observations. For our doctor visits analyses, 885–914 TPS pre-K treatment observations were matched with 467–483 unique comparison observations, and 230–245 CAP Head Start treatment observations were matched with 160–165 unique comparison observations. Finally, for our dental visits analyses, 877–903 TPS pre-K treatment observations were matched with 461–475 unique comparison observations, and 231–271 CAP Head Start treatment observations were matched with 159–166 unique comparison observations.

A key measure of the success of the propensity score matching process is the extent to which the matched treatment and comparison groups have similar observable characteristics. In our analyses of social-emotional outcomes, across five imputations, propensity score matching eliminated almost all statistically significant differences between the matched treatment and comparison groups in 20 individual variables, with a few exceptions. For our analysis of social-emotional phenotypes, TPS pre-K alumni in one imputation were slightly more likely to have their fathers living at home ($p = 0.08$). For our analysis of the attentiveness index, TPS pre-K alumni in one imputation were slightly less likely to be female ($p = 0.10$), and CAP Head Start alumni in one imputation were slightly more likely to be receiving reduced-price lunch ($p = 0.10$).

Matching proved somewhat more challenging for the health outcomes, perhaps because of the smaller sample sizes, though here too there were relatively few remaining imbalances. For our analysis of health status, TPS pre-K alumni in one imputation were slightly less likely to be receiving free lunch ($p = 0.10$), and CAP Head Start alumni in two imputations were slightly more likely to have a mother with a college degree or higher ($p = 0.08$, $p = 0.07$). For our analysis of doctor visits, TPS pre-K alumni were slightly less likely to be receiving a free lunch ($p = 0.06$) in one imputation; slightly less likely to be receiving a reduced-price lunch ($p = 0.10$) and more likely to be receiving a full-priced lunch ($p = 0.09$) in a second imputation; and slightly less likely to have Internet access at home in a third imputation ($p = 0.08$). For our analysis of dental visits, TPS pre-K alumni were slightly less likely to be receiving a free lunch ($p = 0.10$) in one imputation; slightly more likely to be receiving

a full-priced lunch in a second imputation ($p = 0.09$); and slightly less likely to have Internet access at home in a third imputation ($p = 0.08$). CAP Head Start alumni in one imputation were slightly more likely to have a mother with a college degree or higher ($p = 0.08$).

As another measure of balance, we examined the absolute standardized differences between matched treatment and comparison groups in each analysis. This measure is the absolute value of the difference in the sample means between the matched treatment and comparison groups as a percentage of the square root of their averaged sample variances (Rosenbaum & Rubin, 1985). With few exceptions, these differences were below 10 percent, and where larger differences persisted, all were below 15 percent. Propensity score matching enabled us to design treatment and control groups otherwise very similar in observable characteristics. To illustrate the extent to which we were able to improve the comparability of our treatment and comparison samples, consider that for the TPS health status analysis, matching reduced the mean absolute standardized difference in the first imputation from 8.0 to 2.6 percent. The reduction in bias is even more pronounced for our CAP Head Start health status analysis, in which matching reduced the mean absolute standardized difference from 34.8 to 4.9. Using multivariate regression models to analyze these matched samples, as we did when studying social-emotional effects, further increases our confidence in the results.

Findings

In the findings reported below, we offer four sets of results: (i) program impacts for children who participated in the TPS pre-K program; (ii) program impacts for children who participated in the CAP Head Start program; (iii) program impacts for free lunch-eligible children who participated in the TPS pre-K program; and (iv) program impacts for free-lunch eligible children who participated in the CAP Head Start program. The first set of results is authoritative for all students who participated in the TPS pre-K program, regardless of whether they were poor enough to qualify for a free lunch. The second set of results is likewise authoritative for all students who participated in the CAP Head Start program, the majority of whom (90 percent) are eligible for free lunch. Because Head Start participants are disproportionately disadvantaged, a fairer comparison of pre-K and Head Start program effects involves limiting the pre-K sample to free lunch-eligible children. These comparisons do, however, involve smaller sample sizes, and, while they do compare only low-income children in both programs, we do not have income data and thus cannot be assured that the Head Start (or TPS) free-lunch sample is relatively more economically disadvantaged. Indeed, there are a few remaining differences between the free lunch children who attended Head Start, and those who attended pre-K. In order to provide the most valid possible comparison, the third and fourth sets of results thus include program impacts limited to free lunch-eligible children in each program.

As Table 1 indicates, both early childhood programs have had substantial positive impacts on the cognitive development of young children. However, the performance of the TPS pre-K program is clearly superior to that of Head Start for

Table 1. Cognitive Outcomes (Effect Size), Regression-Discontinuity Design

	Full Sample		Free Lunch	
	TPS Pre-K	CAP Head Start	TPS Pre-K	CAP Head Start
Letter-Word ID	0.98***	0.49***	1.20***	0.42**
Spelling	0.74***	0.33**	0.93***	0.25
Applied problems	0.36***	0.36**	0.36***	0.30**

Note: Results are from a regression-discontinuity model that includes controls for gender, race, age, mother’s education, whether the child lives with his or her father, and Internet access at home. TPS, Tulsa Public Schools; CAP, Community Action Project.

** $p < 0.05$; *** $p < 0.01$.

Table 2. Social–Emotional Impacts (Effect Size), Propensity Score Matched Sample

	Full Sample		Free Lunch	
	TPS Pre-K	CAP Head Start	TPS Pre-K	CAP Head Start
“Disobedient”	0.02	0.15	−0.07	0.13
“Aggressive”	0.04	0.02	0.07	0.02
“Attention-seeking”	−0.09	−0.05	−0.08	−0.11
“Apathetic”	−0.09	−0.01	−0.11	−0.20
“Timid”	−0.15**	−0.27*	−0.21*	−0.17
Attentiveness index	0.19***	0.06	0.26**	0.12

Note: Results are from a teacher fixed effects model that includes controls for gender, race, age, mother’s education, whether the child lives with his or her father, and Internet access at home. Comparison observations are weighted to account for matching with replacement, such that treatment observations receive a weight equal to 1, and comparison observations receive a weight equal to the number of times they were matched. Robust standard errors were adjusted to account for clustering by student. TPS, Tulsa Public Schools; CAP, Community Action Project.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

pre-reading (Letter-Word ID test) and pre-writing (Spelling test) outcomes. The superiority of the TPS pre-K program is especially striking if one considers program impacts for children who qualified for a free lunch.

In Table 2, which focuses on social–emotional development, we see some differences between the two programs. Students who participated in the TPS pre-K program are less timid and more attentive than comparable students who did not participate in the TPS pre-K program. In contrast, for Head Start alumni as a whole, there are no statistically significant differences between program participants and comparable students, with the exception of a marginally significant decline in timidity ($p < 0.10$). If we focus on free lunch-eligible students in particular, TPS pre-K alumni also fare better. Specifically, TPS pre-K alumni are more attentive and marginally less timid than comparable students who have not attended TPS pre-K. In contrast, Head Start alumni are indistinguishable from comparable students who have not experienced Head Start. Overall, participation in the TPS pre-K program has enhanced the attentiveness and reduced the timidity of young children, though not dramatically (and, with regard to timidity, only marginally for free lunch students), while participation in the CAP Head Start program has produced no statistically significant effects beyond a marginal reduction in timidity for the full sample.

In Table 3, which focuses on health effects, the outcomes for children who participated in the Head Start program are better than those for children who

Table 3. Health Impacts (Effect Size), Propensity Score Matched Sample

	Full Sample		Free Lunch	
	TPS Pre-K	CAP Head Start	TPS Pre-K	CAP Head Start
Excellent health	-0.01	0.24**	0.04	0.25*
Last doctor visit (<6 months)	-0.10	0.08	-0.08	0.07
Last dentist visit (<6 months)	0.13*	0.49***	0.14	0.49***

Note: Results are from a linear probability model that includes controls for gender, race, age, mother's education, whether the child lives with his or her father, and Internet access at home. Comparison observations are weighted to account for matching with replacement, such that treatment observations receive a weight equal to 1, and comparison observations receive a weight equal to the number of times they were matched. Robust standard errors were adjusted to account for clustering by student. TPS, Tulsa Public Schools; CAP, Community Action Project.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

participated in the TPS pre-K program. Specifically, children who participated in the Head Start program are more likely to be rated by their parents as having "excellent" health than their control group ($p < 0.05$); Head Start alumni are also more likely to have visited a dentist within the last 6 months ($p < 0.01$). For TPS pre-K alumni, we see only a marginally significant positive impact on dental visits ($p < 0.10$). For free lunch-eligible children, we continue to see positive effects for Head Start program participants—an increase in dental visits ($p < 0.01$) and a marginally significant improvement in parent-rated health status ($p < 0.10$). For free lunch-eligible children who participated in TPS pre-K, we see no statistically significant health effects.

Following the method recommended by Paternoster, Brame, Mazerolle, and Piquero (1998), we conducted *z*-tests for the differences between slopes to compare TPS pre-K and CAP Head Start effects. The results of these tests tend to confirm the findings reported above, although not every "difference" visible to the naked eye turns out to be statistically significant at an acceptable level. For the full sample, the TPS pre-K program produces stronger pre-reading scores ($p < 0.01$) and stronger pre-writing scores ($p < 0.05$) than Head Start. In math, the two programs are equally efficacious. Statistically significant differences are also apparent for free lunch-eligible children in particular. For the full sample and for the free lunch-eligible sample, we see no statistically significant differences in social-emotional development between TPS pre-K alumni and Head Start alumni. Finally, for the full sample, Head Start alumni are marginally more likely to be in "excellent" health, as rated by parents, than TPS pre-K alumni ($p < 0.10$), and Head Start alumni are more likely to have visited a dentist recently ($p < 0.05$). For free lunch-eligible children, Head Start alumni are marginally more likely to have visited a dentist recently ($p < 0.10$).

Discussion

Our strategy in comparing Tulsa's leading early childhood education programs has been to compare each with a highly similar control group, through the use of a regression-discontinuity design or propensity score matching.⁹ This produces effect sizes for the TPS pre-K program and the CAP Head Start program, which can then

be compared. We then extend our findings by focusing on free lunch-eligible students in particular.

Although both programs excel at educating young children (cognitive effects), the TPS pre-K program's effects are stronger, for both pre-reading and pre-writing skills. For pre-math skills, Head Start and TPS programs are equally effective. A simple explanation for the early literacy advantages of the school-based program is that they are in the education business, which places a strong emphasis on pre-reading and writing skills, in particular. In contrast, Head Start has multiple missions, which could dilute its educational impact. It is, however, unclear why the Tulsa Head Start program would be at a relative disadvantage to the TPS programs with regard to early literacy education, for which there has been a strong push for program improvement, but not with regard to early math learning. Elsewhere, we have reported that the TPS pre-K program and CAP Head Start differ in their allocation of classroom time. More specifically, TPS pre-K teachers devote more time to math activities, and, at a marginal level, to writing activities, while Head Start teachers devote more time to social studies activities, which included fantasy play (Phillips et al., 2009). Greater time devoted to math and writing and less time devoted to fantasy play helps to explain the TPS advantage in improving pre-reading and pre-writing skills (Gormley et al., 2008). This is not to downplay the potential value of fantasy play, if properly structured. Under the right circumstances, fantasy play can help children to plan and to control their impulses (Brown, 2009). However, substantial allocations of time to fantasy play may come at the expense of other valuable forms of learning.

TPS pre-K teachers also devote more classroom time to math instruction than CAP Head Start teachers, and yet that does not translate into better math outcomes. Perhaps children learn math more through well-chosen examples than through repetition, in which case time on task could matter less for math than for language instruction. It is also the case that, on average, CAP Head Start teachers had 2.5 math courses as undergraduates, as opposed to 1.9 math courses for TPS pre-K teachers ($p < 0.10$), which could translate into higher-quality math instruction despite less time spent on this domain of learning.

It is important to point out that Head Start students are more disadvantaged than TPS pre-K students. Yet if we focus just on free lunch-eligible students, differences in pre-reading and pre-writing (but not pre-math) effects between the two programs persist. In fact, they become even wider, which is consistent with the intuition that disadvantaged children generally benefit more from a high-quality early childhood education program than other children. Some differences in demographic characteristics between TPS pre-K alumni and CAP Head Start alumni do persist, even if we focus on free lunch-eligible children. In particular, Head Start alumni are less likely to be white, more likely to be Hispanic, and less likely to have Internet access at home than TPS pre-K alumni. There are, however, no differences in mother's education, percent African-American, or the presence of a biological father in the home if we focus on free lunch children. The ramifications of these remaining low-income sample differences in the two programs for program impact are, as of yet, unknown. In any event, it is important to stress that we compare each treatment group

(TPS alumni, Head Start alumni) with a tailor-made control group, before we compare the effect sizes of the two treatment groups.

Unlike the early literacy effects, where both programs succeed, the TPS program succeeds more dramatically on pre-literacy outcomes, social-emotional effects are closer to a draw, with a small advantage to children who participated in the TPS program, specifically on attentiveness, and, to a lesser extent, on timidity. This lack of strong evidence for a programmatic advantage on social-emotional development, for which the majority of outcomes showed no significant differences (disobedience, aggressiveness, attention-seeking, and apathy), is not surprising in light of classroom observation evidence that both TPS pre-K and Head Start classrooms in Tulsa were characterized by high levels of emotional support (Phillips et al., 2009).

The evidence regarding attentiveness, however, is important in light of prior evidence that this dimension of social-emotional development predicts school performance (Duncan et al., 2007). One possible explanation for the school-based program's modest advantage on attentiveness and timidity is that 68 percent of TPS pre-K alumni (65 percent if one focuses on free lunch-eligible students) attend the same school as kindergarteners that they attended as preschoolers. By way of contrast, only 19 percent of CAP Head Start alumni (18 percent if one focuses on free lunch-eligible students) attend a kindergarten located at the same school site where they participated in Head Start. This continuity for TPS pre-K alumni could be comforting and reassuring. Even if they attend a different school, the rituals of kindergarten life may resemble those of pre-K more than they resemble those of Head Start. If so, prior exposure to an authentic school environment would be a plus. Indeed, Magnuson et al. (2007) found that children who attended pre-K and kindergarten in the same school did not display the negative social-emotional outcomes found for other children in their analyses of ECLS-K data.

Another part of the puzzle may be the early literacy advantage of the school-based pre-K program, noted above. Because TPS pre-K alumni are better prepared academically than Head Start alumni, they may experience less anxiety and thus greater attentiveness and less timidity in kindergarten classrooms. Thus, positive literacy effects could have positive spillover effects in the social-emotional realm.

Once again, it is useful to ask whether the modest TPS advantage disappears when one focuses exclusively on free lunch-eligible children. In fact, the attentiveness results remain exactly the same, while the timidity results change only slightly, with somewhat less impressive findings for both programs. Also, recall that the Head Start alumni are being compared with a highly similar control group, through propensity score matching and a teacher fixed effects model, before we compare program impacts for TPS and Head Start. Comparisons between TPS and Head Start alumni are indirect, mediated through well-chosen control groups for both.

The most striking evidence of health care impacts is on the provision of dental care. For CAP Head Start, program participation has a positive, statistically significant effect on the child's likelihood of having visited a dentist during the previous 6 months; for TPS pre-K, there is a marginally significant positive impact. In the case of Head Start, the positive impacts are easy to understand. The CAP of Tulsa County

Head Start program provides dental screenings at all sites twice a year. If these screenings, by a dental hygienist, identify a cavity or another dental problem, then the child is referred to a dentist. Because the overwhelming majority of Head Start enrollees are eligible for Medicaid, these costs are typically covered by the government, which makes it relatively easy for parents to act on the referral and take their child to a dentist.

The presence of marginally significant positive effects for TPS pre-K program participants is also understandable. Like CAP Head Start, TPS provides dental screening services and referrals, though not at every school. In general, dental screening services (and referrals leading to actual dentist visits) are more likely to be provided at schools with substantial populations of disadvantaged children. These children, at least, should benefit directly from on-site dental services. In fact, if we break our TPS sample into two (schools that provide dental screening through an organization known as Ocean Dental and schools that do not), we see statistically significant positive effects on dentist visits for the former group (effect size = 0.23, $p < 0.05$), no effects for the latter group.

Overall student health, as perceived by parents, is another area where Head Start appears to have a genuine comparative advantage. All children enrolled in Head Start get a developmental screening within the first 45 days after enrollment and a physical screening within the first 90 days after enrollment. The overwhelming majority of Head Start children also receive a free lunch, whose nutritional value may exceed what they would otherwise have received. These experiences may enhance the overall health of Head Start program participants more than the presence of other sick children damages it. The absence of effects of TPS pre-K program participation on overall health status suggests that the school system's more modest efforts to promote student health may be canceled out by the presence of other children with tummy aches and runny noses.

Contrary to expectations, Head Start does not seem to increase the frequency of doctor visits. Conceivably, the effects of CAP Head Start participation on doctor visits have changed since the time of our study. Beginning in the fall of 2006, CAP Head Start established health clinics at four program sites. This could contribute to more frequent doctor visits in the future. Surprisingly, participation in the TPS pre-K program has a marginally significant negative effect on doctor visits. Perhaps this is a function of question wording: How long has it been since your child's last visit to a clinic, health center, hospital, doctor's office, or other place for routine health care? If pre-K children were required to visit a doctor in the fall of their pre-K year in order to fulfill immunization requirements, then they may have been less likely to visit a doctor the following spring or summer, which is the time period captured by the question. Since the time of our study, TPS established health clinics at a limited number of schools, usually those with more disadvantaged student bodies. Thus, we could see more frequent doctor visits by TPS pre-K students in the future.

If we focus on free lunch-eligible children, Head Start's superior performance in health care remains apparent. Head Start's positive effects on dental visits remain strong, while the TPS effects are no longer marginally significant. Head Start's

positive effects on health status are now marginally significant. For free lunch-eligible children, as for children generally, Head Start performs better on health care than TPS.

Conclusion

Our comparison of the effects of two high-quality early childhood education programs in the same city indicates that the school-based program is more successful in promoting early literacy learning, while Head Start is more successful in health effects. The school system's strong emphasis on education gives it a comparative advantage on early literacy outcomes (but not on math outcomes), while Head Start's multiple missions give it a comparative advantage on health.

The biggest surprise is social-emotional effects. In general, despite Head Start's holistic approach and its oft-stated interest in the whole child, it did not show a comparative advantage in the social-emotional realm. Indeed, to the extent that any significant results emerged, it was the TPS programs that had the advantage, but only clearly for one—attentiveness—of our six outcome measures. A key explanation for this may lie in the advantage seen in early literacy learning for the children in the TPS classrooms, which could reduce anxiety and thus boost their capacity to pay attention in a new environment. Another explanation is that for many students, the environment is not new. They are literally in the same school, which may enhance their comfort level with pay-offs in their engagement in learning.

The policy implications of these findings suggest that the current landscape of early childhood programs, in which Head Start and state pre-K program exist alongside each other, warrants reconsideration. This conclusion is drawn from a single site study, and, indeed, from a site in which both the Head Start and TPS programs are of atypically high quality compared with their counterparts in other states (Phillips et al., 2009). With this caveat in mind, our results do suggest that state governments with high-quality, school-based pre-K and Head Start programs should consider them as having important lessons to learn from each other. Head Start can benefit from the stronger early literacy environments and outcomes, which appear to extend to the children's attentiveness in the classroom, which derive from the TPS programs. The TPS programs, in turn, can learn from the health advantages of the Head Start program. That both programs display benefits in early math learning, and in dental care under some circumstances, and neither program showed detrimental impacts on social-emotional behavior suggests that both can play an important and productive role in the lives of young children and their families.

State-funded pre-K programs are generally less costly than Head Start programs (Besharov, Myers, & Morrow, 2007), which creates a burden of proof for Head Start programs to exhibit advantages that extend beyond those of state-pre-K programs. However, the Tulsa model of establishing a collaborative early education enterprise that embraces both Head Start and school-based pre-K programs holds the promise of creating the circumstances where cross-fertilization across program auspices can occur relatively easily.

The coexistence of two high-quality early childhood programs in the same city—one sponsored by the public schools, one sponsored by Head Start—may work to the benefit of both, as friendly competition encourages both programs to excel rather than to lose ground to the other program. Certainly, the CAP of Tulsa County Head Start program's relatively high salaries are a direct response to the relatively high salaries of the Tulsa Public Schools. The presence of two high-quality programs with somewhat different strengths may also be advantageous to children. Parents of children with precarious health may prefer Head Start, while parents of children with lagging literacy skills may prefer the public schools, given the comparative advantages of the two programs.

Looking specifically at the CAP Head Start program, it is important to stress some significant accomplishments, including tangible gains in pre-reading, pre-writing; comparable gains in pre-math skills; better dental care; and improvements in overall child health as perceived by parents. What is missing here, as in the Head Start Impact Study, is any evidence of more than marginally significant social-emotional improvements for 4-year-olds. We need to understand why this is the case, even in a relatively high-quality Head Start program, given the explicit commitment to this realm of outcomes by Head Start administrators. Perhaps very-low-income children pose distinctive challenges to Head Start teachers. The possible contribution of the peer composition of Head Start classrooms also warrants careful study. In any event, our results indicate that Head Start could benefit from the same focused attention on children's social-emotional development that has been paid in recent years to their pre-learning skills. They also point to the pressing need to generate the additional information needed so that states can make informed judgments about how best to expand different early childhood programs.

William T. Gormley Jr. is University Professor in the Georgetown Public Policy Institute, Georgetown University.

Deborah Phillips is a Professor in the Psychology Department, Georgetown University.

Shirley Adelstein is a Ph.D. student in the Government Department, Georgetown University.

Catherine Shaw is an Environmental Protection Program Manager in the Federal Bureau of Investigation.

Notes

This research was made possible by generous grants from the Foundation for Child Development, the David and Lucile Packard Foundation, the Spencer Foundation, and the A.L. Mailman Family Foundation. The authors would like to thank Carolyn Hill, Ron Zimmer, Lynn Karoly, and Jennifer Brooks, and three anonymous reviewers for helpful comments and Emily Page for valuable research support. The authors would also like to thank the teachers and administrators of the Tulsa Public Schools and the Community Action Project of Tulsa County for their cooperation.

1. Tested kindergarten students were somewhat less likely to be poor and black and somewhat more likely to be middle-class and white than the universe of kindergarten students.
2. Assessed kindergarten students were somewhat less likely to be poor and somewhat more likely to be middle-class than the universe of kindergarten students.

3. Assessed kindergarten students were somewhat less likely to be poor and black and somewhat more likely to be middle-class and white than the universe of kindergarten students.
4. We excluded students who were either too young or too old to be in Tulsa pre-K, Head Start, or kindergarten based on their birth dates.
5. This should be less of a problem for dental visits. However, the frequency of doctors' visits might also confound relationships between early childhood education program participation and the child's overall health.
6. Results from a linear probability model differ very little from logistic regression results. Furthermore, this approach is justified by the distribution of the outcome variables, all of which have means near 0.5.
7. In contrast, each teacher filled out approximately two dozen ASPI survey forms.
8. The covariates include gender, age, age², age³, race/ethnicity, eligibility for a free or reduced price lunch, mother's education, Internet access, the number of books at home, the primary language spoken by the child at home, whether the child participated in day care as a 3-year-old, whether the child participated in Head Start as a 3-year-old, whether the child participated in preschool as a 3-year-old, whether the child lives with his/her biological father, foreign-born parents, the mother's marital status, missing data dummies, and a variety of interaction terms. The covariates differed slightly for each sub-sample in order to maximize the balancing of observables.
9. As noted earlier, we also utilized a teacher fixed effects model with covariates for our analysis of social-emotional outcomes but not for our analysis of health outcomes.

References

- Barnett, W. Steven, Dale Epstein, Allison Friedman, Judi Boyd, and Jason Hustedt. 2008. *The State of Preschool 2008*. New Brunswick, NJ: National Institute for Early Education Research.
- Besharov, Douglas, Justus Myers, and Jeffrey Morrow. 2007. "Costs Per Child for Early Childhood Education and Care: Comparing Head Start, CCDF Child Care, and Prekindergarten/Preschool Programs." Washington, DC: American Enterprise Institute, June 13.
- Brown, Emma. 2009. "Debate on Playtime's Value Grows." *Washington Post*, November 21.
- Campbell, Frances A., Craig T. Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson. 2002. "Early Childhood Education: Young Adult Outcomes from the Abecedarian Project." *Applied Developmental Science* 6 (1): 42–57.
- Croy, Calvin, and Douglas Novins. 2005. "Methods for Addressing Missing Data in Psychiatric and Developmental Research." *Journal of the American Academy of Child & Adolescent Psychiatry* 44 (12): 1230–40.
- Currie, Janet, and Duncan Thomas. 1995. "Does Head Start Make A Difference?" *American Economic Review* 85 (3): 341–64.
- . 1999. "Does Head Start Help Hispanic Children?" *Journal of Public Economics* 74 (2): 235–62.
- Dehejia, Rajeev H., and Sadek Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94 (448): 1053–62.
- . 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics* 84 (1): 151–61.
- Duncan, Greg, Amy Claessens, Aletha Huston, Linda Pagani, Mimi Engel, Holly Sexton, Chantelle Dowsett, Katherine Magnuson, Pamela Klebanov, Leon Feinstein, Jeanne Brooks-Gunn, and Kathryn Duckworth. 2007. "School Readiness and Later Achievement." *Developmental Psychology* 43 (6): 1428–46.
- Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. "Longer Term Effects of Head Start." *American Economic Review* 92 (4): 999–1012.
- Gormley, William. 2007. "Early Childhood Care and Education: Lessons and Puzzles." *Journal of Policy Analysis and Management* 26 (3): 651–89.
- . 2008. "The Effects of Oklahoma's Pre-K Program on Hispanic Children." *Social Science Quarterly* 89 (4): 916–36.

- Gormley, William, and Ted Gayer. 2005. "Promoting School Readiness in Oklahoma." *Journal of Human Resources* 40 (3): 533–58.
- Gormley, William, Ted Gayer, Deborah Phillips, and Brittany Dawson. 2005. "The Effects of Universal Pre-K on Cognitive Development." *Developmental Psychology* 41 (6): 872–84.
- Gormley, William, Deborah Phillips, and Ted Gayer. 2008. "Preschool Programs Can Boost School Readiness." *Science* 320: 1723–24.
- Gormley, William T., Deborah A. Phillips, Katie Newmark, Kate Welti, and Shirley Adelstein. 2011. "The Social-Emotional Effects of Early Childhood Education Programs in Tulsa, OK." To appear in *Child Development*.
- Henry, Gary, Bentley Ponder, Dana Rickman, Andrew Mashburn, Laura Henderson, and Craig Gordon. 2004. "An Evaluation of the Implementation of Georgia's Pre-K Program: Report of the Findings from the Georgia Early Childhood Study (2002–03)." Atlanta: Georgia State University, December.
- Kaplan, David W., Claire D. Brindis, Stephanie L. Pibbs, Paul Melinkovich, Kelly Naylor, and Karin Ahlstrand. 1999. "A Comparison Study of an Elementary School-Based Health Center: Effects on Health Care Access and Use." *Archives of Pediatrics and Adolescent Medicine* 153 (3): 235–43.
- Kisker, Ellen E., and Randall S. Brown. 1996. "Do School-Based Health Centers Improve Adolescents' Access to Health Care, Health Status, and Risk-Taking Behavior?" *Journal of Adolescent Health* 18 (5): 335–45.
- Leuven, Edwin, and Barbara Sianesi. 2003. "PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing." <http://ideas.repec.org/c/boc/bocode/s432001.html>. Accessed July 12, 2010.
- Loeb, Susanna, Margaret Bridges, Daphna Bassok, Bruce Fuller, and Russell Rumberger. 2006. "How Much Is Too Much? The Influence of Preschool Centers on Children's Social and Cognitive Development." *Economics of Education Review* 26 (1): 52–66.
- Lowenstein, Amy. 2009. *Fostering the Socio-Emotional Adjustment of Low-Income Children: The Effects of Universal Pre-Kindergarten and Head Start in Oklahoma*. Unpublished doctoral dissertation, Georgetown University, Washington, DC.
- Ludwig, Jens, and Douglas L. Miller. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression-Discontinuity Design." *Quarterly Journal of Economics* 122 (1): 158–208.
- Ludwig, Jens, and Deborah Phillips. 2008. "Long-Term Effects of Head Start on Low-Income Children." *Annals of the New York Academy of Sciences* 1136: 257–68.
- Lurie, Nicole, Ellen Jones Bauer, and Cynthia Brady. 2001. "Asthma Outcomes at an Inner-City School-Based Health Center." *Journal of School Health* 71 (1): 9–16.
- Magnuson, Katherine, Christopher Ruhm, and Jane Waldfogel. 2007. "Does Prekindergarten Improve Preparation and Performance?" *Economics of Education Review* 26 (1): 33–51.
- Paternoster, Raymond, Robert Brame, Paul Mazerolle, and Alex Piquero. 1998. "Using the Correct Statistical Test for the Equality of Regression Coefficients." *Criminology* 36 (4): 859–66.
- Phillips, Deborah, William Gormley, and Amy Lowenstein. 2009. "Inside the Pre-K Classroom Door." *Early Childhood Research Quarterly* 24 (3): 213–28.
- Puma, Michael, Steven Bell, Ronna Cook, and Camilla Heid. 2010. *Head Start Impact Study: Final Report*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Puma, Michael, Steven Bell, Ronna Cook, Camilla Heid, and Michael Lopez. 2005. *Head Start Impact Study: First Year Findings*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Reynolds, Arthur, and Judy Temple. 1995. "Quasi-Experimental Estimates of the Effects of a Preschool Intervention." *Evaluation Review* 19 (4): 347–73.
- Reynolds, Arthur J., Judy A. Temple, Dylan L. Robertson, and Emily A. Mann. 2001. "Long-Term Effects of an Early Childhood Intervention on Educational Achievement and Juvenile Arrest." *Journal of the American Medical Association* 285 (18): 2339–46.

- Richmond, Julius, Deborah Stipek, and Edward Zigler. 1979. "A Decade of Head Start." In *Project Head Start: A Legacy of the War on Poverty*, ed. Edward Zigler, and Jeanette Valentine. New York: Free Press, 135–52.
- Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.
- . 1985. "Constructing A Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *American Statistician* 39 (1): 33–38.
- Royston, Patrick. 2004. "Multiple Imputation of Missing Values." *The Stata Journal* 4 (3): 227–41.
- . 2005a. "Multiple Imputation of Missing Values: Update." *The Stata Journal* 5 (2): 1–14.
- . 2005b. "Multiple Imputation of Missing Values: Update of Ice." *The Stata Journal* 5 (4): 527–36.
- Rubin, Donald. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- . 1996. "Multiple Imputation after 18+ Years." *Journal of the American Statistical Association* 91 (434): 473–89.
- Santelli, John, Anthony Kouzis, and Susan Newcomer. 1996. "School-Based Health Centers and Adolescent Use of Primary Care and Hospital Care." *Journal of Adolescent Health* 19 (4): 267–75.
- Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, William S. Barnett, Clive R. Belfield, and Milagros Nores. 2005. *Life-Time Effects: The High/Scope Perry Preschool Study Through Age 40*. Ypsilanti, MI: High/Scope Press.
- Silberberg, Mina, and Joel Cantor. 2008. "Making the Case for School-Based Health: Where Do We Stand?" *Journal of Health Politics, Policy and Law* 33 (1): 3–37.
- Sinharay, Sandip, Hal Stern, and Daniel Russell. 2001. "The Use of Multiple Imputation for the Analysis of Missing Data." *Psychological Methods* 6 (4): 317–29.
- U.S. Department of Health and Human Services. 2005. *Head Start Impact Study: First Year Findings*. Washington, DC: Author.
- U.S. Department of Health and Human Services, Administration for Children and Families. 2008. http://www.acf.hhs.gov/programs/ohs/about/index.html#prog_desc. Accessed July 13, 2010.
- Webber, Mayris P., Kelly E. Carpinello, Tosan Oruwariye, Yungtai Lo, William B. Burton, and David K. Appel. 2003. "Burden of Asthma in Inner-City Elementary Schoolchildren. Do School-Based Health Centers Make A Difference?" *Archives of Pediatrics and Adolescent Medicine* 157 (2): 125–29.
- Wong, Vivian, Thomas Cook, W. Steven Barnett, and Kwanghee Jung. 2008. "An Effectiveness-Based Evaluation of Five State Pre-K Programs." *Journal of Policy Analysis and Management* 27 (1): 122–54.
- Xiang, Zongping, and Lawrence J. Schweinhart. 2002. "Effects Five Years Later: The Michigan School Readiness Program Evaluation through Age 10." Paper prepared for the Michigan State Board of Education. Ypsilanti, MI: High-Scope Educational Research Foundation.
- Young, Thomas L., Sandra L. D'angelo, and James Davis. 2001. "Impact of A School-Based Health Center on Emergency Department Use by Elementary School Students." *Journal of School Health* 71 (5): 196–98.
- Zigler, Edward, and Susan Muenchow. 1992. *Head Start: The Inside Story of America's Most Successful Educational Experiment*. New York: Basic Books.