

HEAD START'S COMPARATIVE ADVANTAGE:

MYTH OR REALITY?

William T. Gormley, Jr.
Georgetown Public Policy Institute
Georgetown University

Deborah Phillips
Psychology Department
Georgetown University

Shirley Adelstein
Government Department
Georgetown University

Catherine Shaw
Federal Bureau of Investigation

Paper presented at the Annual Meeting of the Association for Public Policy
Analysis and Management (APPAM), Washington, D.C., November 6, 2009.

ABSTRACT

In recent years, Head Start's primacy in early childhood education has been challenged by state-funded pre-K programs, some of which are based in public schools. An analysis of data from Tulsa, Oklahoma indicates that school-based pre-K programs are more effective in improving cognitive and social-emotional outcomes, while Head Start is more effective in improving health outcomes. Head Start's comparative advantage is the delivery of health services, including dental health. It is less efficacious in the other two areas.

HEAD START'S COMPARATIVE ADVANTAGE: MYTH OR REALITY?

For many years, the Head Start program enjoyed a virtual monopoly over early childhood education services for disadvantaged four-year-old children. In recent years, that primacy has been challenged by state-funded pre-K programs, which now serve more four-year-olds than Head Start (Barnett et al. 2008). Most, though not all, of the four-year-olds served by state pre-K programs are disadvantaged.

Are disadvantaged four-year-olds better served by Head Start or by state-funded pre-K programs? The question is of considerable interest to both parents and public officials. From the parents' perspective, Head Start is the better bet only if it produces better educational outcomes or some other combination of positive outcomes for their children. 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.

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 non-educational 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 and Muenchow 1992: 18). The Head Start program’s seven objectives included: 1. improving the child’s physical health and physical abilities; 2. helping the emotional and social development of the child by encouraging self-confidence, spontaneity, curiosity, and self-discipline; 3. improving the child’s mental processes and skills, with particular attention to conceptual and verbal skills; 4. establishing patterns and expectations of success for the child that will create a climate of confidence for future learning efforts; 5. 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; 6. 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 7. increasing the sense of dignity and self-worth within the child and his family (Richmond, Stipek & Zigler 1979: 137).

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 A.A., B.A., or advanced degree in early childhood education or a degree in a related field, with preschool teaching experience.

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 & Human Services 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.

Literature Review

Early intervention (EI) programs for children in poverty seek to compensate for sub-optimal 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 (Schweinhart, et al., 2005; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Reynolds, Temple, Robertson, & Mann, 2001). 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 non-poor 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 (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 (Magnuson et al. 2007; Loeb et al. 2006). State-specific studies have also documented positive effects on language skills and math skills (Henry et al. 2004, Wong et al. 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 and Temple 1995).

The Head Start Impact Study, a randomized experiment funded by the federal government, found evidence of positive, but modest, cognitive impacts for three-year-olds and four-year-olds, though not in every area (Puma et al., 2005). For three-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 four-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 three-year-olds or four-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.

In our research in Tulsa, we have found even stronger impacts for the CAP of Tulsa County Head Start program on children's early literacy and math skills at kindergarten entry (Gormley, Phillips & Gayer, 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 and Thomas 1995; Currie and Thomas 1999; Garces et al. 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 (Magnuson et al. 2007; Loeb et al. 2006), 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. Non-experimental national studies found negative effects of Head Start participation on social-emotional behavior among kindergarteners (Magnuson et al. 2007; Loeb et al. 2006).

In our own work in Tulsa (Gormley, Phillips, Newmark, & Perper, in press; Lowenstein, 2009), participation in the TPS pre-K program was associated with lower teacher ratings on timidity, attention-seeking behavior, and apathy, 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 sample 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 three-year olds (the effect size was .12), but no statistically significant effect for four-year-olds (Puma et al. 2005). Effects on dental care were more striking, with effect sizes of .34 for three-year-olds and .32 for four-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 four-year-olds; 69 percent of the treatment group had dental case, as opposed to 52 percent of the

control group, among three-year olds (Puma et al., 2005, p. xv). Ludwig and Miller (2007), focusing on some of the earliest Head Start participants, found that a 50-100% 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% to 50% of the control mean.

Evidence on the effects of school-based health care programs is more limited but growing. A recent study in Newark, N.J. 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 and Cantor 2008). In contrast, other studies have found that school-based clinics do reduce hospitalizations and visits to the emergency room (Santelli et al. 1996; Kaplan et al. 1999; Young et al. 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 et al. 2001; Webber et al. 2001). 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 and 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

Data

The data available to us, never before discussed in one article, come from four sources: cognitive testing conducted in Tulsa, Oklahoma 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.

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.

Methodology

The greatest challenge 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; Gormley, Phillips & Gayer 2008). We have used a similar strategy to estimate the impacts of the Tulsa Head Start program on children's cognitive development (Gormley, Phillips & Gayer 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 four-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 pre-test. 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, Phillips, Newmark, and Perper 2009). 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 and has the additional advantage of controlling for clustering by classroom.

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's visits, then it becomes difficult to disentangle doctor's visits due to immunization schedules from doctor's 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 treatment 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 and Wahba 1996; Dehejia and Wahba 2002; Rosenbaum and Rubin 1983; Rosenbaum and Rubin 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 and 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. An identical procedure 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 .001 for the estimation of TPS pre-K treatment effects and .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 (.001) and Head Start (.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 analyses, 1,114 TPS pre-K treatment observations (85 percent) were matched with 558 unique control observations, and 300 CAP Head Start treatment observations (83 percent) were matched with 191 unique control observations. For our health analyses, matched sample sizes varied slightly across outcome variables, because in some cases parents responded to some but not all of the three health-related survey items. For each of the three TPS pre-K health outcomes, approximately 81 percent of the treatment sample was matched (n=857 for health status; n=846 for doctors' visits; and n=838 for dentists' visits). The total number of unique control observations that were used for these analyses included 438 observations for health status, 432 observations for doctors' visits, and 427 observations for dentists' visits. For each of the three CAP Head Start health outcomes, approximately 72 percent of the treatment sample was matched (n=195 for health status and doctors' visits, and n=192 for dentists' visits). The total number of unique control observations that were used for these analyses included 134 observations for health status and doctors' visits and 132 observations for dentists' visits. In each analysis, some control observations were matched with multiple treatment observations due to matching with replacement.

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 and health outcomes, propensity score matching eliminated almost all significant differences. There were only three exceptions. First, for social-emotional outcomes, TPS pre-K alumni were slightly more likely to have internet at home than members of the matched control group. Second, for the TPS pre-K health status analysis, TPS pre-K alumni were less likely to be Hispanic than members of the matched control group. Finally, for the CAP Head Start doctors' visits analysis, CAP Head Start alumni were more likely to have paid lunch status than members of the matched control group. Propensity score matching enabled us to design treatment and control groups otherwise very similar in observable characteristics. 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: 1) program impacts for children who participated in the TPS pre-K program; 2) program impacts for children who participated in the CAP Head Start program; 3) program impacts for free lunch-eligible children who participated in the TPS pre-K program; and 4) 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. In order to provide the most valid possible comparison, the third and fourth set 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 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 apathetic, less timid, and less likely to engage in attention-seeking behavior than comparable students. They are also more attentive. In contrast, for Head Start alumni as a whole, there are no statistically significant differences between program participants and comparable students. If we focus on free lunch-eligible students in particular, TPS pre-K alumni also fare better. Specifically, TPS pre-K alumni are less disobedient, less timid, and more attentive than comparable students. In contrast, Head Start alumni are largely indistinguishable from comparable students, though they are marginally more likely to be aggressive ($p < .10$). Overall, participation in the TPS pre-K

program has enhanced the social-emotional development of young children, whereas participation in the CAP Head Start program has not.

In Table 3, which focuses on health effects, the outcomes for children who participated in the Head Start program are somewhat better than those for children who 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. Both sets of children – TPS pre-K alumni and CAP Head Start alumni – are more likely to have visited a dentist within the last six months. For free lunch-eligible children, program impacts are more similar and limited to dental visits. However, it is worth noting that the effect size for Head Start program participation is larger than that for TPS pre-K program participation, as it was for children as a whole.

Following the method recommended by Paternoster et al. (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 basically confirm the findings reported above, though 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 ($p = .01$) and stronger pre-writing scores ($p < .05$) than Head Start. Statistically significant differences are also apparent for free-lunch eligible children in particular. For the full sample, TPS pre-K alumni are more attentive than Head Start alumni ($p < .10$). For the free-lunch eligible sample, TPS pre-K alumni are less disobedient than the Head Start alumni ($p < .05$). Finally, for the full sample, Head Start alumni are more likely to be in “excellent” health than TPS pre-K alumni ($p < .05$), and

Head Start alumni are also more likely to have visited a dentist recently ($p < .05$). For free-lunch eligible children, differences in program effects are not statistically significant.

Discussion

Our strategy in comparing Tulsa's leading early childhood education programs has been to compare each to 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 also test the robustness of 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, there are no differences in program efficacy. A simple explanation for the educational advantages of the school-based program is that they are in the education business. In contrast, Head Start has multiple missions, which could dilute its educational impact. 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 academic subjects, while Head Start teachers devote more time to fantasy play (Phillips, Gormley, & Lowenstein, 2009). Greater time devoted to letters and sounds and less time devoted to fantasy play help to explain the TPS advantage in improving pre-reading and pre-writing skills (Gormley et al. 2008).

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 ($p < .10$), which could translate into higher quality math instruction despite less time spent on this domain of learning.

It is certainly fair 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 cognitive effects between the two programs persist. In fact, they become even wider. Thus, the educational advantages of the school-based program cannot be explained away by arguing that Head Start teachers are trying to educate less educable 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. But 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 cognitive effects, where both programs succeed but the TPS program succeeds more dramatically on pre-literacy outcomes, social-emotional effects are marked by positive outcomes for TPS, but no effects (or negligible effects) for Head Start. One might imagine that the two programs differ in the level of emotional support

provided by classroom teachers, but that is not the case. In fact, classroom observations indicate high levels of emotional support for both TPS pre-K and CAP Head Start classrooms (Phillips, Gormley, & Lowenstein, 2009).

One possible explanation for the school-based program's social-emotional advantage is that approximately two-thirds of TPS pre-K alumni attend the same school as kindergarteners that they attended as preschoolers. This 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. By way of contrast, only 19 percent of CAP Head Start alumni attend a kindergarten located at the same school site where they participated in Head Start.

Another critical part of the puzzle is the cognitive 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 exhibit greater confidence and self-assurance in kindergarten classrooms. Thus, positive cognitive effects could have positive spillover effects in the social-emotional realm. Better preparation for the academic side of school may reduce timidity and apathy and enhance attentiveness.

Once again, it is useful to ask whether the TPS advantage diminishes or disappears when one focuses exclusively on free-lunch eligible children. Clearly, that is

not the case. Although the number of statistically significant positive effects declines from four to three for TPS pre-K program participation, we now see for the first time a statistically significant negative effect for Head Start – namely, increased aggressiveness. Thus, the school-based program’s advantage in generating positive social-emotional outcomes cannot be attributed to a higher percentage of free-lunch eligible students in Head Start. Also, recall that the Head Start alumni are being compared to 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 both TPS pre-K and CAP Head Start, program participation has a positive, statistically significant effect on the child’s likelihood of having visited a dentist during the previous six months. 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 statistically 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.

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 doctors' visits. Conceivably, the effects of CAP Head Start participation on doctors' 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 doctors' visits in the future. Surprisingly, participation in the TPS pre-K program has a marginally significant negative effect on doctors' 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 doctors' 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 diminishes. There are no longer statistically significant differences in the effects of program participation on overall health, and the curious marginally significant negative effect of TPS on doctors' visits disappears. However, differences in dental care remain, with Head Start continuing to perform better 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 two of the three areas (cognitive effects, social-emotional effects), while Head Start is somewhat more successful in the third area (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. Despite Head Start's holistic approach and its oft-stated interest in the whole child, the school system winds up producing better social-emotional effects. A key explanation for this may lie in the superior educational accomplishments of the school system, which reduce anxiety and

boost self-confidence 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 and reduce acting-out behaviors.

The policy implications of these findings are not obvious. These conclusions are 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 to their counterparts in other states. With this caveat in mind, our results do suggest that state governments with high-quality, school-based pre-K programs should consider them as a highly viable option – and perhaps as an alternative to Head Start -- for preparing 4-year old children for the academic and social demands of formal schooling. State-funded pre-K programs are generally less costly than Head Start programs (Besharov, Myers, and Morrow 2007), and yet they produce impressive cognitive benefits and modest social-emotional improvements. They even seem to improve dental care. With more on-site health services, they might improve general student health as well.

As for Head Start, 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 statistically significant social-emotional improvements for four-year-olds. We need to understand better why this is not happening, despite a commitment to this outcome 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.

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

			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: * = p-value < 0.10; ** = p-value < 0.05; *** = p-value < 0.01

Table 2. Social-emotional Impacts (Effect Size), Propensity Score Matching + Teacher Fixed Effects Model

Social-emotional Factor	TPS Pre-K	CAP Head Start	Free Lunch	
			TPS Pre-K	CAP Head Start
"Disobedient"	0.00	0.05	-0.16**	0.17
"Aggressive"	0.00	0.18	0.07	0.22*
"Attention-Seeking"	-0.12**	0.00	-0.07	-0.03
"Apathetic"	-0.15**	-0.02	-0.02	-0.08
"Timid"	-0.13**	-0.05	-0.21***	-0.11
Attentiveness Index	0.19***	-0.04	0.25***	0.07

Note: * = p-value < 0.10; ** = p-value < 0.05; *** = p-value < 0.01

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

			Free Lunch	
	TPS Pre-K	CAP Head Start	TPS Pre-K	CAP Head Start
Excellent Health	-0.05	0.24**	0.12	0.12
Last Doctor Visit (<6 mo)	-0.10*	0.06	-0.02	0.20
Last Dentist Visit (<6 mo)	0.21***	0.48***	0.27***	0.50***

Note: * = p-value < 0.10; ** = p-value < 0.05; *** = p-value < 0.01

REFERENCES

- Barnett, W. Steven, Dale Epstein, Allison Friedman, Judi Boyd, and Jason Hustedt. 2008. *The State of Preschool 2008*. New Brunswick, N.J.: 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, D.C.: American Enterprise Institute, June 13.
- 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, pp. 42-57.
- Currie, Janet and Duncan Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review* 85, pp. 341-64.
- Currie, Janet and Duncan Thomas. 1999. "Does Head Start Help Hispanic Children?" *Journal of Public Economics* 74, pp. 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 (December), pp. 1053-1062.
- Dehejia, Rajeev H. and Sadek Wahba. 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies," *Review of Economics and Statistics* 84 (February), pp. 151-161.
- 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, pp. 1428-46.
- Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. "Longer Term Effects of Head Start," *American Economic Review* 92, pp. 999-1012.
- Gormley, William and Ted Gayer. 2005. "Promoting School Readiness in Oklahoma," *Journal of Human Resources* 40 (Summer), pp. 533-58.
- Gormley, William, Ted Gayer, Deborah Phillips, and Brittany Dawson. 2005. "The Effects of Universal Pre-K on Cognitive Development," *Developmental Psychology* 41 (December), pp. 872-84.
- Gormley, William, Deborah Phillips, and Ted Gayer. 2008. "Preschool Programs Can Boost School Readiness," *Science* 320 (June 27), pp. 1723-24.

Gormley, William, Deborah Phillips, Katie Newmark, and Kate Perper. 2009. "Socio-Emotional Effects of Early Childhood Education Programs in Tulsa." Paper presented at the Society for Research in Child Development Conference, Denver, Colo., April 3.

Gormley, William T., Deborah A. Phillips, Katie Newmark, and Kate Perper. In Press. "The Social-Emotional Effects of Early Childhood Education Programs in Tulsa, OK," *Child Development*.

Gormley, William. 2007. "Early Childhood Care and Education: Lessons and Puzzles," *Journal of Policy Analysis and Management* 26 (June), pp. 651-89.

Gormley, William. 2008. "The Effects of Oklahoma's Pre-K Program on Hispanic Children," *Social Science Quarterly* 89 (December), pp. 916-36.

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. Phibbs, Paul Melinkovich, Kelly Naylor, and Karin Ahstrand. 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, pp. 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, pp. 335-345.

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>.

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, pp. 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, D.C.

Ludwig, Jens and Miller, Douglas L. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression-Discontinuity Design," *Quarterly Journal of Economics* 122, pp. 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*, pp. 1-12.
- Ludwig, Jens and Deborah Phillips. 2007. "The Benefits and Costs of Head Start," *Social Policy Report XXI*(3), pp. 1-18.
- Lurie, Nicole, Ellen Jones Bauer, and Cynthia Brady. 2001. "Asthma Outcomes at an Inner-City School-Based Health Center," *Journal of School Health* 71, pp. 9-16.
- Magnuson, Katherine, Christopher Ruhm, and Jane Waldfogel. 2007. "Does Prekindergarten Improve Preparation and Performance?" *Economics of Education Review* 26, pp. 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, pp. 859-66.
- Phillips, Deborah, William Gormley, and Amy Lowenstein. 2009. "Inside the Pre-K Classroom Door," *Early Childhood Research Quarterly*.
- Puma, Michael, Steven Bell, Ronna Cook, Camilla Heid, and Michael Lopez. 2005. "Head Start Impact Study: First Year Findings." Washington, D.C. 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, pp. 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, pp. 2339-2346.
- Richmond, Julius, Deborah Stipek, and Edward Zigler. 1979. "A Decade of Head Start." In Edward Zigler and Jeanette Valentine, *Project Head Start: A Legacy of the War on Poverty*. New York: Free Press, pp. 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 (April), pp. 41-55.
- Rosenbaum, Paul R. and Donald B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *American Statistician* 39 (February), pp. 33-38.
- 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, pp. 267-275.

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 (February), pp. 3-37.

U.S. Department of Health and Human Services. 2005. "Head Start Impact Study: First Year Findings." Washington, DC: Author.

U.S. Department of Health & Human Services, Administration for Children & Families. 2008. http://www.acf.hhs.gov/programs/ohs/about/index.html#prog_desc.

Webber, Mayris P, Kelly E. Carpintello, 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, pp. 125-129.

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* (January), pp. 122-54.

Xiang, Zongping & Schweinhart, Lawrence J. 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), pp. 196-198.

Zigler, Edward and Susan Muenchow. 1992. *Head Start: The Inside Story of America's Most Successful Educational Experiment*. New York: Basic Books.

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 three-year-old, whether the child participated in Head Start as a three-year-old, whether the child participated in preschool as a three-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.