GETTING INSIDE THE "BLACK BOX" OF HEAD START QUALITY:
WHAT MATTERS AND WHAT DOESN'T?

Janet Currie
Matthew Neidell

Getting Inside the "Black Box" of Head Start Quality: What Matters and What Doesn't?
Janet Currie and Matthew Neidell
NBER Working Paper No. 10091
November 2003
JEL No. I28, I38

ABSTRACT

Critics of Head Start contend that many programs spend too much money on programs extraneous to education.   On the other hand, Head Start advocates argue that severely disadvantaged children need a broad range of services. Given the available evidence, it has been impossible to assess the validity of these claims.   In this study, we match detailed administrative data with data on child outcomes from the National Longitudinal Survey of Youth 1979, including test scores, behavior problems, and grade repetition. We find that former Head Start children have higher reading scores and are less likely to have been retained in grade where Head Start spending was higher.   Holding per capita expenditures constant, children in programs that devoted higher shares of their budgets to education and health have fewer behavior problems and are less likely to have been retained in grade.   However, when we examine specific educational inputs holding per capita expenditures constant, only pupil/teacher ratios matter.

Janet Currie
Department of Economics
UCLA
405 Hilgard Avenua
Los Angeles, CA 90095-1477
and NBER
currie@simba.sscnet.ucla.edu

Matthew Neidell
Department of Economics and CISES
University of Chicago
5734 S. Ellis Avenue
Chicago, IL 60637
mneidell@uchicago.edu

Introduction

Head Start is a preschool program for disadvantaged children that aims to provide early intervention so that children can begin schooling on an equal footing with their more advantaged peers. The program is designed to address a wide variety of needs in addition to providing early educational experiences. For example, federal guidelines mandate that children receive nutritious meals, that their medical needs be assessed, and that parents be involved in the program. Begun in 1965 as part of President Johnson's "War on Poverty", in 2000 Head Start served 857,664 children in predominantly part-day programs, about 65% of eligible 3 and 4 year old poor children.[1] Over time, federal funding has increased from $96 million in 1965 to $5.3 billion in 2000.

There have been dozens of studies of Head Start and related preschool and early school enrichment programs.[2] These studies have established that early intervention programs can have a dramatic effect on children's lives. For example, the Perry Preschool Project and the Carolina Abecedarian Projects demonstrated long-term gains in academic achievement, as well as reductions in teen pregnancy and crime, in the context of well-executed randomized trials. Although Head Start Centers are typically of lower quality than these model programs (but of higher average quality than other child care available to low income parents), Head Start has also been shown to have both short-run and long-run positive impacts on children. Garces, Thomas,

------

[1] Authors' calculations based on the 2000 Head Start Fact Sheet at http://www.acf.hhs.gov/programs/hsb/research/factsheets/00_hsfs.htm, and statistics on the number of children less than 6 in poverty at http://ferret.bls.census.gov/macro/032001/pov/new01_003.htm.

[2] See Currie (2001), Barnett (1995) and Karoly et al. (1998) for reviews.

and Currie (2002) find long-term effects of Head Start participation on the educational

achievements and criminal behavior of young adults, while Currie and Thomas (1995) find

higher test scores and lower grade repetition among former child participants. However, Head

Start does not appear to bring poor children up to the average levels of achievement of non-poor

children.

The recognition that Head Start may not be enough to prevent the intergenerational

transmission of poverty has led to divergent prescriptions for reform. On the one hand, critics

argue that Head Start's core mission should be educational, and that in many centers funding for

educational services has been diverted to, for example, programs for parents (Abell Foundation,

2000).   Concern has also been expressed about low levels of teacher pay and qualifications. The

1990 Bush Presidential Campaign's proposal to transfer responsibility for the program from the

Department of Health and Human Services to the Department of Education symbolized this view

that the program should be focused more exclusively on education.

On the other hand, advocates argue that disadvantaged children have a wide range of

needs that must be met before they can learn academic material effectively, so that it is

appropriate for Head Start to offer a range of services.   For example, in coordination with state

agencies, Head Start currently provides disabled children (broadly defined to include children

with developmental delays) with speech and language therapy, occupational and physical

therapy, special education, and mental health services.[3]   Advocates argue that more funding is

---

[3] According to the Head Start performance standards laid out in the Federal Register, a
disabled child is one with mental retardation, hearing or visual impairments, autism, brain injury,
and other health impairments or specific learning disabilities. At state discretion, the term may
also be applied to children suffering developmental delays. States are mandated to provide
services to disabled preschool children, and may choose to do so either through Head Start or in

necessary in order to raise Head Start quality while continuing to provide these services (Zigler and Styfco, 1994).

The Head Start reauthorization bill introduced by Representatives Castle and Boehner on May 22, 2003 took a strong position on the way in which Head Start should be reformed. It included language stressing the importance of academic preparation and focusing on teacher qualifications (requiring that 50 percent of teachers have a bachelor's degree within five years, as well as requiring all new teachers to have at least an associates degree within 3 years). Since the bill provided few new resources for Head Start, money for improving teacher qualifications would have to be taken from other aspects of the program.

These considerations make it especially timely and important to "get inside the black box" of Head Start program quality, and determine which aspects of the program have the greatest impact on child outcomes. There has been little previous research on this question, primarily because of a paucity of data linking Head Start program characteristics to child outcomes.[4] This paper provides a look into the box, by linking administrative data on Head Start program quality to information about child test scores, behavior problems, and grade repetition from a large, national sample of children drawn from the National Longitudinal

conjunction with Head Start. E.g. Head Start may be used to provide an "inclusive experience" for a child who receives more specialized therapy elsewhere.

[4] The most recent federally-sponsored study of Head Start is the Family and Child Experiences Survey (Zill, Resnick and McKey, undated). Unfortunately this study took a short-term perspective, following children over the course of one year in Head Start, and had no control group. The study found that children showed gains in social skills over the course of the year. Cognitive gains were assessed by comparing the Head Start children to national norms. These findings were consistent with those of many other studies which have documented short-term gains, particularly to verbal skills. A large-scale, federally-sponsored, experimental evaluation of Head Start's effects on short-term outcomes is currently in the field.

Survey of Youth 1979 Child and Young Adult Data (NLSY). We estimate models in which the key variable is the interaction between whether or not a child attended Head Start, and per capita spending on Head Start in the year and county that a child was four years old. That is, we examine the difference in future outcomes between Head Start children in high and low spending areas, using the difference between non-Head Start children in high and low spending areas as a control.

We find that relative to other children who were in the same location at age four, Head Start children have higher reading scores and are less likely to have been retained in grade where Head Start spending was higher. Holding per capita expenditures constant, children in programs that devoted higher shares to education and health also do better. However, when we examine specific educational inputs, only pupil/teacher ratios matter. Teacher salaries and the fraction of teachers with qualifications including the BA degree have little effect on child outcomes. These surprising results may reflect the fact that most Head Start teachers already have qualifications, and that salaries are not as low as has been suggested, given that most Head Start teachers work part time. Our findings suggest that current proposals for the reform of Head Start may not have the desired impact.

The rest of the paper is laid out as follows: Section II provides some background information about Head Start, Section III describes the data, section IV discusses methods, results are presented in section V, and section VI offers some conclusions.

II. Background

Head Start is a federal-local matching grant program. Local grantees apply to the

Department of Health and Human Services (DHHS) for funds to run their programs on a three year cycle. Head Start grantees must provide 20 percent of program costs, either in cash or in kind. Incumbents appear to have priority for funding. For example, they have the right to appeal if their applications for funding are turned down, while would be new entrants do not. Applications are evaluated on the basis of: "the extent to which the applicants demonstrate in their application the most effective Head Start program"; the cost-effectiveness of the proposed program; the qualifications and experience of the applicant and of the staff; the capability to adhere to the Head Start Performance Standards; and the need for Head Start services in the community (DHHS, 2001, section 1302.1).

Every year, the federal government appropriates funding for Head Start. This funding is then allocated to each state using a formula that depends on the relative number of children from birth to age four who are living in families with incomes below the poverty line in each state, as compared to all states (DHHS, June 2002). In addition to their regular grant applications, grantees can apply for additional funds for cost of living adjustments, quality improvements, and for Training and Technical Assistance (two percent of all Head Start funds are set aside for this purpose). Money is also made available periodically for special initiatives. For example, in 2002, $6 million was available for initiatives to "develop new and innovative approaches designed to increase the involvement of fathers in their local Head Start program", and DHHS announced that they would fund between 10 and 30 of these special grants in each region. A further $3 million was available for programs that would promote "positive youth development" through involvement in local Head Start programs (DHHS, 2002b). In 2003, special funds were available for Head Start programs that had partnerships with historically black universities,

and/or with Latino service agencies (DHHS, 2003). These special initiatives are advertised in the Federal Register.

These considerations suggest that grantees will receive larger amounts of funding if : the federal government allocates more money for Head Start in the annual appropriation; if grantees have the expertise to write good proposals and stay on top of special program announcements; or if grantees can raise a larger matching component locally, are in a high-need area, and/or have relatively few other programs applying in their state.   In addition, per capita funding could vary because of fluctuations in enrollments or because of inflation given the three year funding cycle. However, we found that enrollment changed relatively little over time in most programs, which is consistent with anecdotal evidence that many programs have waiting lists. And inflation was low over most of the period we consider.

Two of these potential sources of variation in per capita expenditures have the potential to complicate the interpretation of our results.   First, to the extent that good Head Start directors both write better proposals, and run better programs, we could wrongly attribute the effect of having a better director to spending. We will test for this possibility below using data on the qualifications of program directors.   Second, if it is easier to raise matching funds in wealthy areas, and these areas also offer other enrichment activities for children, then we may wrongly attribute the effects of local amenities to Head Start spending.   The inclusion of non-Head Start children as a control group helps to control for this possibility, but we also include detailed controls for characteristics of the child's county of residence at age four and state fixed effects, and we estimate models using only the subset of relatively poor counties.   It is worth noting that to the extent that local areas direct more resources to places with especially disadvantaged

children, we may also end up under-estimating the effects of spending on Head Start.

Finally, it is useful to consider the way that children are selected into Head Start programs. Program standards direct agencies to develop a recruitment process to help them reach "those most in need of Head Start services" and specify that the "use of referrals from other public and private agencies" is one way to achieve this goal (U.S. Dept. of Health and Human Services, 2003, Section 1305.5). Children on the waiting list are to be ranked in terms of each agency's selection criteria to insure that the neediest are admitted first. It will be important to keep this selection in mind in what follows.

III. Data

Data about Head Start programs comes from two sources: Program Information Reports (PIRs) and an administrative data set on Head Start budgets called PCCOST. The PIRs are filled out by each Head Start grantee, delegate agency and Parent Child Center (PCC) in most years, and cover the period 1988 to 2000.[5] They provide information about the educational qualifications of staff, teacher pay, teacher turnover, and much more.[6] The PCCOST data spans the period 1990 to 2001 and has information about the sources and disposition of Head Start program funds, breakdowns of enrollment by the type of children served, and information about

---

[5] As the names suggest, the federal government gives money to a Head Start grantee, which may or may not pass the money onto one or several "delegate agencies". (That is, the grantee could also be the agency that ran the center). A unit of observation in the administrative data is therefore defined by the grantee's number, the delegate agency's number, and the year.

[6] In principal, the administrative data contains information about many aspects of Head Start programs. However, in practice, data for many of the variables is often coded as missing or zero (when it could not be zero). In this paper, we focus on the budget data, which is relatively complete, and on selected inputs which do not suffer as greatly from missing data problems.

the director and other administrative staff. The PCCOST data also has the FIPS county code, which enables us to merge it with the NLSY data.

The PCCOST data was generated by a computer program that many agencies used to fulfill their administrative reporting requirements. However, agencies were not required to use the program, and some did not do so in every year. Moreover, the program was phased in during the early 1990s, and started to be phased out around 1998 (in favor of a replacement program) so that the years 1996, 1997, and 1998 are those with the most complete reporting. Hence, in order to preserve sample size, we went through the data casting backward and forward as many as three years to fill in missing observations. The administrative data is also fairly noisy, necessitating some data cleaning and removal of outliers. For example, we deleted a small number of agency-year observations that reported enrollments of fewer than 10 children, or per capita expenditures of less than $500 per child.

The NLSY data tracks the children of the roughly 6,000 young women who took part in the original NLSY survey, which began in 1979. Hence, a great deal of information is available about the characteristics of mothers, including their county of residence in each year, and their score on the Armed Forces Qualifications Test (AFQT), a test of job skills. Beginning in 1986, the children have been surveyed and assessed biannually. Mothers were asked whether the child had ever attended Head Start and whether the child had repeated any grades.[7] For each child,

---

[7] The grade repetition questions were only asked to children 10 and over in the 1986 and 1988 surveys. Hence, there are fewer observations for grade repetition than for some of the other outcomes. Also, in 2000, the survey did not ask about whether Kindergarten had been repeated. In addition to the question about Head Start, mothers were also asked whether children attended some other form of preschool, which should have helped to clarify that Head Start is distinct from other preschools. In our data, 69 percent of children are reported to have attended other preschools, so that some children attended both Head Start and other preschools.

we have scores for tests of vocabulary (the Peabody Picture Vocabulary Test, or PPVT), reading (the Peabody International Assessment Tests for reading recognition (PIAT-RR) and reading comprehension (PIAT-RC)), mathematics (PIAT mathematics), as well as an index of behavior problems (normalized separately by sex).[8]

There are different numbers of observations available for each test, given that they were administered to children in different age ranges. Also, the reading comprehension test was administered only to children who scored above a threshold level on the reading recognition test. We use all non-missing observations for each score, in order to preserve the maximum possible number of observations.[9]

It is important to examine a wide range of indicators, particularly since there is a good deal of controversy about the use of standardized tests to measure student achievement.   Among the test scores we examine, minority children perform particularly badly on the PPVT, and better (relative to white non-Hispanic children) on the PIATs.   There is increasing evidence that non-cognitive skills are important in determining eventual outcomes (c.f. Heckman et al., 2000). For example, evidence from evaluations of model preschool programs suggests that there may be improvements in long-term outcomes including educational attainment even in the absence of lasting gains in test scores. Hence, we examine grade repetition and behavior problems.

In order to merge the administrative data with the NLSY data, we proceeded as follows. First, we identified the calendar year in which a child was four years old.   Second, we retained

---

[8] The NLSY offers two normalizations of the Behavior Problems Index. Results using the index that is not normalized by sex produced very similar results.

[9] For further information about these data, see the NLSY79 Child and Young Adults User's Guide 2000 which is available at http://www.bls.gov/nls/y79cyaguide/nlsy79cusg.htm.

scores only for children 60 months of age and older. In this way, we avoid using scores that could have been measured before the child attended Head Start.   Third, we took the mean of all age-normalized percentile scores available for each child, in order to arrive at one observation per child.[10] One advantage of this procedure relative to examining test scores at a particular age, is that there are many fewer missing values. For example, a child who was surveyed at age 4 in 1990, skipped an interview, and then was reassessed at age 8 in 1994 would not have any test score available for age 6. A second advantage is that averaging over several test scores is likely to give a more accurate measure of the child's abilities than taking a single test score. Previous analyses of these data have shown that as much as half of the total variation in test scores is within child, rather than between child (Currie and Thomas, 1995).

We lose 403 children for whom information on Head Start attendance is missing, 292 children with poor matches to the administrative data (as discussed further below), 179 children for whom information on the mother's AFQT scores was missing, and we also exclude a further 111 children whose reported permanent income (defined as the mean over all of the incomes reported by the mother in the NLSY) was greater than $150,000, for a total usable sample of 4,468 children.

Three complications arose with the actual merge of the two data sets.   First, while most counties were served by only one Head Start program, some counties are served by more than

---

[10] On average there are 2.8 observations per child on the PIAT-MATH and PIAT-RR tests, 2.3 observations on the PIAT-RC, and 1.4 observations on the PPVT. Thus, we might expect estimates for PPVT to be noisier than those the PIATs.

one program.[11] In these cases, we took a weighted average of the characteristics of programs

serving the county, where the weights were the number of children in each program. Second,

even after filling in missing values in the way described above, there were instances in which we

could not find an appropriate administrative match for the NLSY data in a particular year.   In

these cases, we took data from the nearest available year.   We kept only NLSY observations that

could be matched to an administrative data point within plus or minus three years of the year that

the child was aged 4. Third, we assume that all of the NLSY children attended center-based

programs, though a small number of them could have been in primarily home-based programs.

The implicit assumption we have relied on in filling in missing administrative data, is

that programs change relatively slowly over time within counties. In fact, if we decompose the

variance in the administrative data into within and between components, we generally find that

the lion's share of the variation is across agencies. For example, the between component of the

variance in enrollments is 449 compared to a within component of 108, while the comparable

figures for per capita expenditures are 1682 vs. 827.   By smoothing within-agency variation in

spending, we will tend to attenuate its effects, and thus the estimates presented below may be

under-estimates of the true effects of Head Start expenditures on outcomes.

Table 1 shows means of the administrative data for all agency-years with information on

per capita expenditures, for agency-years that are missing data on per capita expenditures (and

thus are excluded from our analysis), and for the subset of the data that we match to NLSY

children. The NLSY data are further broken out by race and ethnicity, as well as by whether the

---

[11] The average number of programs per county was 1.5, but the 50th percentile of the
distribution of the number of programs was 1, the 75th percentile was 2, and the 95th percentile
was 3.

county was poor, as discussed further below. All dollar amounts are in real 1998 dollars.

The first three rows of Table 1 show total enrollments, per capita funding "on the balance sheet", and total per capita funding including "in-kind" transfers.[12] Head Start centers are required to give breakdowns of the "on the balance sheet" portion of their budgets into ten categories including: administration, education, services for the disabled, occupancy costs (i.e. rent, utilities, etc.), health, parent services, social programming, nutrition programs, transportation and "other", so we will focus most of our investigation on this on-the-balance sheet portion of the budget, though we show below that our results are not sensitive to the choice of expenditure measure.   The shares spent on various services are viewed as an important indicator of program priorities. For example, the PCCOST manual states that "a grantee with 80% of its budget allocated to Education would clearly be stating that providing educational services was of paramount importance to that agency" (DHHS, 1999, page 27). We aggregate expenditures related to disability, nutrition, and health into one category, so that we are left with expenditures on education, all health, and "other".

We include information about the race and ethnic breakdown of children served by the programs[13], as well as information about some specific educational inputs that are often

---

[12] Cash on the balance sheet includes money from the Administration for Children, Youth, and Families, and "non-federal share cash". The total spending also includes a small amount of "other cash", money from USDA, "non-federal share" in-kind, and "other in-kind". One reason for keeping track of the budget in these categories is that programs are required to partially match federal contributions with cash and in-kind funding from other sources. There are also limits on how much can be spent on particular budget items. For example, programs cannot spend more than 15 percent of on the balance sheet funds on administration.

[13] It is important to keep in mind that the Hispanic sample in the NLSY includes only children born to mothers who were already in the country in 1978. Hence, it is not representative of the experience of more recent arrivals, or of immigrant children.

examined in the literature on school quality. These include pupil/teacher ratios, teacher salaries, the fraction of teachers with qualifications, the average education of Head Start teachers, and the qualifications of the Head Start program director.

Table 1 shows first, that there is little difference in other respects between agency-years for which per capita funding data was reported, and agency-years for which it is missing. Columns 3 and 4 break out programs with high and low per capita spending, in order to see whether there are any systematic differences in spending patterns across these programs. While there is considerable variation across all programs in, for example, the share of the budget spent on health related activities, the share does not differ systematically between high and low spending programs. Rather, the mean shares spent on different activities are remarkably similar across the columns, suggesting that higher spending programs spend proportionately more on all types of activities.

Table 1 also shows that relative to all programs, the programs matched to NLSY children are larger (mean enrollments of 704 vs. 487) and funded at somewhat lower levels (total per capita funding of $4793 vs. $5260). We believe that these differences reflect the sampling scheme in the NLSY–sampled children were apparently unlikely to be located in counties that had very small Head Start programs. The NLSY children also tend to be in programs that have lower fractions of white children, and higher fractions of black and Hispanic children, on average. Despite these differences, other aspects of the programs are quite similar–the main exceptions are that compared to the full sample, the NLSY children have teachers and directors

who are slightly better paid, and have higher pupil/teacher ratios.[14]

The Table indicates that despite concerns about teacher qualifications, 85 percent of teachers are qualified. The 1994 Head Start Reauthorization Act required that by September 30, 1996, Head Start classroom teachers had to have a Child Development Associate degree; a state-awarded certificate for preschool teachers; an associate, baccalaureate, or advanced degree in early childhood education; or a degree in a field related to early childhood, with experience in teaching and a state awarded certificate to teach in a preschool program. We do see an increase in the fraction of teachers with qualifications in our data, from 80 percent in 1991/92 to 92 percent in 1997. Twenty-two percent of programs report that their average teacher has a B.A. degree or higher, a fraction that had increased to 34 percent by 1997.

It is striking that despite their qualifications, teachers receive what seems be very low pay, averaging $16,428 per year. It is possible that this reflects the part-time nature of the typical program. The data indicate that on average, Head Start teachers work 5 hours per day, for 193 days during the year. Hence, this salary implies an hourly wage rate of $17.50 per hour, which is comparable to the average hourly wage of $17.93 ($1998) that workers with B.A. degrees received in 2001 (authors' computation from the May Current Population Survey). The

---

[14] Note that the pupil/teacher ratios in Table 1 are higher than the mandated maximum class size. Head Start classes are supposed to have between 15 and 20 students. The reason for this discrepancy is that we are dividing all children who have been enrolled by the number of teachers, which tends to give a high estimate. That is, to the extent that there is turnover (on average children in our sample were in programs with 7.7% turnover), more children will have been enrolled in the program than are present at any point in time. We chose this estimate because it is difficult to know exactly how many children are enrolled at a point in time, and for consistency with the per capita expenditure figures which are computed using all children served by the program in the fiscal year. We have also estimated models similar to those in Table 7 using the interaction of turnover rates with Head Start as our "input" and did not find any statistically significant effect.

average salary received by Head Start program directors is $36,877 which is similar to the salary of a typical worker with a BA.[15]

Per capita funding levels in Table 1 can be compared to those for "model" programs. The part-day Perry Preschool intervention cost $12,884 per child (in 1999 dollars) for a program that lasted eight months a year over two years. Since 20 percent of the children participated only for one year, the figures imply that the cost per child was approximately $7,000 per year, so that Head Start costs about 71 percent of what Perry Preschool cost (Karoly et al., 1998). The preschool component of the Carolina Abecedarian project cost about $15,000 per child, per year, and this part of the intervention lasted five years.[16]

The last four columns of Table 1 explore differences in the programs serving children in poor counties, and children of different racial/ethnic backgrounds. The children's location is measured as of the year that they were 4. Children are said to be in a poor county if the fraction of families in poverty in 1989 was greater than 11 percent in their county. Eleven percent was the median percent poor in our sample. Programs in poor counties, and those serving black and Hispanic children tend to be much larger, and are slightly worse funded on average than those serving white children: Average total funding per child is $4,894 for white children, $4,611 for Hispanic children, $4,748 for black children, and $4,619 in poor counties. This finding is

---

[15] These salary figures do not include fringe benefits, and Head Start teachers may or may not receive such benefits as health insurance and retirement plans. Note, that the PCCOST data attributes spending on benefits to categories of spending in the proportions that personnel are allocated to these categories. Thus, spending on benefits is accounted for in the budget shares.

[16] Fewell and Scott (1999) report that the Infant Health and Development Program, another well-known early intervention whose long term effects are currently being assessed, also cost about $15,000 per year per child.

consistent with Resnick and Zill's (undated) findings regarding the lower quality of programs with higher shares of minority children. The share of the budget spent on different categories is quite similar across the four columns, with the exception that programs in poor counties and those serving minority children tend to place greater emphasis on educational expenditures.

One of the most striking differences between programs is in the composition of peers. White children attend programs with the highest fraction white (49 percent), while black children attend programs that are 66 percent black and Hispanic children attend programs that are 45 percent Hispanic.

Table 2 shows the fraction of the NLSY children who were ever enrolled in Head Start, measures of child outcomes, and maternal AFQT (one of the most important indicators of family background and predictors of child test scores).   Twenty-three percent of the children were reported to have ever attended Head Start. In recent years, about 800,000 have been enrolled in every year, which works out to about 14 percent of all 4 year old children and 7 percent of all three year old children being served at any point in time (ACYF, 1998). Given that children can attend Head Start over a two year window, the NLSY figures seem reasonably consistent with the administrative data. Participation rates are also much higher for black than for white children, which again is consistent with administrative data.[17]

The test scores that we use are normalized so that the child's score is the percentile of the national distribution for children of the same age. By this metric, we can see that NLSY children

---

[17] In 1998, the fractions of black, Hispanic, and white Head Start children were 36, 26, and 31 percent respectively–if all poor children had participated at equal rates, the corresponding fractions would have been 29, 31, and 35 (authors' calculation based on the assumption that poverty rates for white, black, and Hispanic children were 15, 40, and 40 percent, respectively.

have average scores on the PIAT-Math, somewhat higher scores on the PIAT-Reading Recognition and Reading Comprehension tests, and quite low scores on the PPVT. Given the sampling frame of the NLSY, it tends to oversample children born to younger mothers, so other things being equal, one might expect NLSY children to have lower rather than higher than average scores. The observed pattern may be because the norms used for the PIAT tests are relatively dated. The NLSY children also have scores on the behavior problems index that are higher than the national norm, indicating a greater incidence of behavior problems. There are striking differences in patterns of these scores across race and ethnic groups, with whites typically having better scores than blacks and Hispanics. Children in poor counties, also tend to score more poorly. Finally, 12.3 percent of children had ever repeated a grade by the time of the 2000 survey, and this rate varies from a low of 8.6 percent among whites to a high of 18 percent among black children.

Differences in mothers' backgrounds show typical patterns across race and ethnicity. White mothers have much higher AFQT scores than other mothers (50.97, 19.89, and 24.43, for white, black, and Hispanic mothers respectively). White mothers also have permanent incomes (calculated as the mean over all reported incomes in the NLSY sample) of $51,240 compared to $30,610 for black mothers. Children of black and Hispanic mothers also have more siblings than children of white mothers, and have mothers with less education on average though these differences are not shown in the Table.

The second panel of Table 2 shows the differences between the Head Start children and other children within each group. Head Start children have significantly worse outcomes in every way than other children within the same group. For example, the average Head Start child

has a PIAT-RR score 8.7 points lower than the average child, a difference of approximately 1/3 of a standard deviation. The last two columns show, not surprisingly, that Head Start children have mothers with much lower AFQT scores and permanent incomes than other children.

Panel 3 show mean differences between Head Start and other children, adjusting for the observable characteristics of children and mothers that we include in our regression models, which are described further below. A comparison of Panel 2 and Panel 3 suggests that some but not all of the difference between Head Start and other children can be attributed to differences in observable background characteristics. For example, differences in PIAT-MATH scores become statistically insignificant, and the gap in PIAT-RR scores is reduced to 2.6 points. The fact that some differences between Head Start and other children remain is not surprising, given the way that these children are selected.

IV. Methods

We estimate least squares regression models of the following form:

(1) $\text{Outcome}_i = a + a_1\text{Percap}_c + a_2\text{HeadStart}_i + a_3\text{Percap}_c*\text{HeadStart}_i + a_4X_i + a_5Z_c + a_6S + a_7\text{Cohort}_i + e_i$,

where Outcome is a test score, the behavior problem index, or a measure of whether the child has repeated a grade; i indexes the individual and c indexes the county. Percap is per capita expenditure in the county's Head Start program; Head Start is an indicator equal to one if the child attended Head Start and zero otherwise; X is a vector of mother and child characteristics

18

including: AFQT, permanent income, mother's education (dropout, high school, some college, college), mother's race/ethnicity, mother's number of siblings, child's gender, whether the child is the first born, and whether the child has more than five siblings; Z is a vector of county characteristics including population in 1990, the percent of the population that was black in 1990, the percentage of the population that was Hispanic in 1990, the percent of births to teenage mothers in 1988, median family income in 1989, the percentage of families below poverty in 1989, the percent of votes cast for the Democratic Presidential candidate in 1992, the percent of the population under 18 years in 1990, and the percent older than 64 years in 1990; and S is a vector of state fixed effects.[18]

As we saw in Table 2, this set of variables explains a considerable portion of the difference between Head Start and other children. The inclusion of the county-level controls, and of the state dummy variables is intended to control for factors that might be correlated both with variation in per capita expenditures on Head Start and with child outcomes. Moreover, as discussed above, per capita funding may vary with the fraction of poor children in the state, which could also be related to child outcomes. Cohort is a vector of dummy variables for the year that the child was aged 4 (1988-89, 1990-91, 1992-93, 1994-95, 1996-97, 1998-99), which allows outcomes to vary for children of different cohorts. Finally, e is an idiosyncratic error term.

---

[18] Several previous studies compare siblings in the NLSY in order to identify the effect of Head Start on outcomes (c.f. Currie and Thomas, 1995; Currie and Thomas, 2000; Garces, Thomas, and Currie, 2002). The focus in this study is different–we wish to ask whether, conditional on having attended Head Start, the size of any estimated effect on outcomes is affected by per capita spending on Head Start? For our purposes, a sibling comparison would be less than ideal, given the imprecise nature of the matching between the NLSY and the administrative data.

Percap, the main effect of per capita spending, captures the overall effect of being in a high spending county, for both Head Start and non-Head Start children. Counties with high spending on Head Start could have unobserved characteristics that are associated with higher test scores for all children. For example, some counties might have generous community health programs in addition to Head Start. To the extent that counties with these characteristics also spend more on Head Start, conditional on the included state and county characteristics, the effects will be captured by Percap. Table 2 suggests that the coefficient on HeadStart, a_2, is likely to be negative because children who attend Head Start are negatively selected relative to other children. Thus, we can view Head Start as a proxy for the negative unobserved characteristics of children that are associated with attending Head Start.

Our main focus is on a_3, the coefficient on the interaction between Percap and Head Start. This coefficient measures the effect of additional spending on children who attended Head Start. We are essentially examining the difference between Head Start children in high and low spending areas, using the differences in outcomes between non-Head Start children in high and low spending areas as a control. We might find for example, that Head Start children did better in high spending areas than in low spending areas, but if non-Head Start children showed a similar pattern, then a_3 would not be statistically significant, and we would have to attribute the better results in the high spending areas to some other characteristic of those areas.

A potential problem with this research strategy is that the Head Start children may be selected differently in high and low spending counties. Suppose, for example, that high spending programs attract more able students. Then it will appear that spending improves outcomes, whereas in reality, it only changes the way that children are selected. In addition to

the extensive set of state and county level controls included in our regressions, we have adopted

three strategies to deal with this potential problem. First, we conduct all of our analyses

separately for children who were in counties with high poverty rates at age 4. Our hypothesis is

that the difference between Head Start and non-Head Start children will be more uniform within

this set of counties than it is in the country as a whole. As we will see, our results tend to be

stronger in this more homogeneous set of counties, suggesting that heterogeneity in the way that

Head Start students are selected across counties is not driving our results.

Second, we directly examine the way that spending affects the selection of Head Start

children in terms of their mother's AFQT scores, which is one of the most powerful predictors of

the children's test scores. We will show that higher spending programs do tend to attract

children with slightly better maternal AFQT scores, but the change in the way that children are

selected in terms of this observable factor is far too small to account for the gains that we find.

Third, we conduct a similar analysis to see if higher spending programs are less likely to

select children of low birth weight, defined as birth weight less than 2500 grams. Children of

low birth weight are more likely to have a range of physical and behavioral problems than other

children, so if higher spending programs selected fewer low birth weight children, then this

could have an effect on measured outcomes. However, we find no evidence of selection of this

type. Thus, if selection on unobservables follows a similar pattern to selection on AFQT and

birth weight, it is unlikely that such selection could explain our results.

We are also interested in assessing the extent to which different types of expenditures

have different effects on outcomes. In order to address this issue, we estimate models in which

we control not only for per capita expenditures, but also for the fraction of the budget spent on

education, and on all health-related activities defined as the sum of the budget shares spent on health care, nutrition, and services for the disabled. As Table 1 indicates, these expenditures together account for approximately 11 percent of Head Start "cash on the balance sheet" expenditures.

The model we estimate takes the following form:

(2) $Outcome_i = a + a_1 Percap_c + a_2 HeadStart_i + a_3 Percap_c * HeadStart_i + a_4 ShareEd_c + a_5 ShareEd_c * HeadStart_i + a_6 ShareHealth_c + a_7 ShareHealth_c * HeadStart_i + a_8 X_i + a_9 Z_c + a_{10} S + a_{11} Cohort_i + v_i,$

Finally, we investigate the effects of particular educational inputs, including pupil-teacher ratios, pupil-classroom staff ratios, average teacher salaries, the fraction of teachers who have qualifications, education of the average teacher, and Head Start director's qualifications. Specifically, we estimate models of the form:

(3) $Outcome_i = a + a_1 Percap_c + a_2 HeadStart_i + a_3 Percap_c * HeadStart_i + a_4 Input_c + a_5 Input_c * HeadStart_i + a_6 X_i + a_7 Z_c + a_8 S + a_9 Cohort_i + v_i,$

where Input is one of the educational inputs and the other variables are defined as described above. This specification holds per capita spending constant. Hence, it asks what would happen if we increased one input holding the total budget constant, which is analogous to what legislation requiring increased spending on particular inputs without increasing overall budgets would do.

In addition to estimating these models separately for children in poor counties, we

estimate them separately for children of different race/ethnic groups. The fact that children from these groups attend largely separate programs, have very different mean outcomes, and have systematically different backgrounds suggests that the effects of Head Start could differ between groups.

V. Results

a) Effects of Head Start Spending on Child Outcomes

Estimates of the effects of per capita spending in Head Start on child outcomes are shown in Table 3. The measure of spending used here is cash on the balance sheet.   As discussed above, the main effect of per capita spending captures the influence of unobservables associated with having lived in a particular county in the child's last possible Head Start year.   These effects are not statistically significant, suggesting that the included regressors do a good job of controlling for other factors that are associated with both per capita spending and child outcomes.

The estimated main effects of Head Start are not statistically significant for PIAT-MATH, behavior problems, or grade repetition. However, they are significantly negative for the reading and vocabulary scores (PIAT-RR, PIAT-RC, and PPVT). The point estimates of -8.9, -8.7, and -7.8 imply that children who went to Head Start would have scores that were a third of a standard deviation lower than those of other, in the absence of the program.   At the mean level of spending of $3,500, the coefficients imply a gap between Head Start and other children of 2.8 points for PIAT-RR, which is very similar to what we saw in Panel 3 of Table 2. The significant interactions for these reading and vocabulary scores imply that an increase in spending of $1000

23

would reduce gaps in verbal scores by between 1.4 and 1.7 points, so that an increase in spending of $1,600 (i.e. total spending of $5,100 "on the balance sheet") would eliminate the gap, if the effects were linear.

Of the variables included in our models, AFQT is among the most important determinants of child outcomes. The point estimates in Table 3 imply that the child's test scores rise by about a quarter point for every one point gain in maternal AFQT, while the probability of grade repetition falls. These magnitudes will be useful in assessing the effects of expenditure-related selection into the program below.

Table 4 presents estimates obtained from several subsamples of the NLSY data. Panel 1 shows estimates for children in poor counties. Although the sample size has been cut in half, the estimated effects of the interaction between Head Start and spending rises for reading scores (PIAT-RR and PIAT-RC). The point estimate on PPVT is very similar to that obtained in the full sample, although the standard error rises. Thus, the results appear to hold within this sample of relatively homogeneous counties, as well as in the full sample.

Panels 2 through 4 show separate estimates for children of each race/ethnic group. The results are very striking. The point estimates for the effects of expenditures on PIAT-RR and PPVT scores among white children are similar to those reported above, although they are not statistically significant. However, spending does appear to have a significant negative effect on the probability of grade repetition: Every $1,000 increase in spending is associated with a four percentage point reduction in the probability that a white Head Start child repeated a grade. The main effects of Head Start suggest that white children who attended Head Start have test scores similar to those of other white children once observable differences are controlled for, but that

they are still 19.7 percentage points more likely to have repeated a grade in the absence of the program, so the improvement associated with increased spending is significant.

Among black children, higher expenditures appear to increase test scores across the board, but have little effect on the probability of grade repetition. It is particularly notable that spending appears to increase math scores as well as reading and vocabulary scores. This pattern of results is consistent with the fact that the main effects of Head Start suggest that, conditional on observables, black children who attended Head Start would have had lower test scores in the absence of the program, but the same probability of grade repetition.

We find no significant effects of expenditures among Hispanic children, though this could be due in part to the relatively small sample size. Currie and Thomas (1999) found large positive effects of Head Start attendance in the NLSY's Hispanic children. Hence, an alternative interpretation of the current finding is that while attendance has a positive effect, additional spending does not.

The last panel of Table 4 gives results using total per capita spending rather than "cash on the balance sheet". We focused on the "on the balance sheet" measure first because this measure can be broken out by type of expenditure, and second, because all federal payments are included in the cash on the balance sheet measure. Table 4 establishes that the results are qualitatively similar for both measures, though the point estimates are smaller. In fact, the relative magnitudes are consistent with the interpretation that, on average, only spending "on the balance sheet" affects child outcomes.

The first panel of Table 5 presents estimates of the "effects" of spending on maternal AFQT. These regression models are of the same form as (1) except that AFQT is the

dependent variable, and they omit maternal AFQT, permanent income, and maternal education from the right hand side "X" vector. Clearly, spending can have no causal impact on AFQT since in most cases it was measured before the child was born, so any significant effect represents a change in the way that children are selected into the program. The first two columns indicate that the interaction between per capita spending and "Ever Head Start" is not statistically significant in either the whole sample, or in the poor county sample. However, it does appear that if we estimate the sample separately by racial and ethnic groups, high spending programs attract white and black children with slightly higher maternal AFQT scores than lower spending programs, while there is no evidence of such selection among Hispanics.

Among whites, each $1000 increase in per capita expenditure is associated with 2.5 point reduction in the maternal AFQT gap between Head Start and other children. Table 4 allows us to say whether this is a large or small effect: It indicates that each 1 point increase in AFQT was associated with a one tenth of a percentage point reduction in the probability of grade repetition, while Head Start was associated with a four percentage point reduction in the probability of grade repetition. That is, it would take an increase of 40 AFQT points to equal the effect of Head Start, although the mean level of AFQT in the sample of whites is only 51. Hence, these estimates suggest that differential selection of children with better backgrounds into high spending programs is unlikely to account for much of the estimated effect of Head Start.

Among blacks, a similar comparison suggests that differential selection is likely to explain only a small part of the estimated effect. In addition, it is important to note that observable differences in AFQT are controlled for in the Table 4 models, so that selection on this important observable indicator does not bias the estimated coefficients.

26

The second panel of Table 5 shows estimates for models in which the dependent variable is whether or not a child was low birth weight. Again, this is something that could not be influenced by the program, so any effect would represent selection into the program. However, the estimates indicate that there is no relationship between spending and whether a child who attended was low birth weight.

Although we cannot rule out the possibility that Head Start children are selected very differently in terms of unobservables in high spending and low spending programs, the fact that selection on AFQT and birth weight is either negligible or non-existence suggests that this is unlikely.

In summary, these estimates indicate that Head Start programs that spend more per capita have larger positive effects on children's reading and vocabulary scores. These effects are more pronounced when we focus only on children in high poverty counties. When we look at racial and ethnic subgroups, we find that among white children higher spending results in less grade repetition for Head Start children. Among black children, higher spending increases Head Start math, reading, and vocabulary test scores, but has no effect on grade repetition.

b) Getting inside the Black Box of Program Design

The results discussed thus far suggest, in contrast to much of the literature on primary and secondary education, that increasing per capita spending levels on Head Start might have positive effects in terms of promoting educational attainment. However, there are many different ways that additional spending could be allocated. As discussed above, some observers feel that Head Start should focus on the educational aspect of its mandate, spending more money on

qualified teachers, and less money on other "extraneous" programs. Other observers feel

strongly that, particularly for preschool children, it is important to offer a comprehensive

package of services. In this section, we ask whether it is possible to shed light on this debate

using data about the way that Head Start programs allocate their budgets.

Table 6 shows estimates from models which control both for per capita expenditures and

for the *share* of the budget spent on education and health (i.e. the aggregate of the health,

nutrition, and disability portions of the budget).[19] The first panel of the table suggests that

allocating spending to health increases reading scores, while increasing the share of the budget

devoted to education reduces behavior problems and grade repetition. The second panel

confirms these results in the subsample of children from high poverty counties. Both sets of

effects are quite large. For example, the second panel point estimate of 63.6 on the interaction of

the share of health and "ever Head Start" in the equation for PIAT-RR, implies that a 5 percent

increase in the share of spending devoted to health would increase the reading scores of Head

Start children (relative to other children) by 3.2 points. Similarly, a 5 percent increase in the

share of the budget devoted to education would result in almost a 1.3 point reduction in the score

on the behavior problems index among the children in poor counties, and in a 1.5 percentage

point reduction in the probability of grade repetition.

Hence given the "adjusted" mean differences between Head Start and other children in

poor areas shown in Table 2, these estimates imply that shifting five percent of the budget

---

[19] As discussed above, higher spending programs tend to spend more on all components so that there is a good deal of collinearity between the levels of spending on different components of the program. On the other hand, there is no reason that the share of spending on health or education should vary systematically with per capita expenditures.

towards health (holding per capita expenditures and the share of education constant) would eliminate the gap in PIAT-RR, and that shifting five percent of the budget towards education (holding per capita expenditures and the share of health constant) would close half of the gap in the behavior problems index, and one fifth of the gap in the probability of grade repetition between Head Start and other children.

The third panel of Table 6 shows results using the share spent on education combined with the share spent on health, nutrition, and disabilities. This share can be thought of as the portion of the budget devoted directly to spending on children. The estimates are similar to those in Panel 2, in that they suggest that increasing the share of spending devoted to children, holding per capita expenditures constant would reduce behavior problems and the probability of grade repetition.

We have also estimated these models separately for the three racial/ethnic groups but the standard errors are large, so that it is difficult to say whether the shares actually have differential effects on the three groups. The estimates suggest however, that increasing the share spent on education reduced behavior problems among blacks, and reduced the probability of grade repetition among Hispanics.

The administrative data allow further investigation of the question of whether any specific educational inputs can be shown to be beneficial. For example, the emphasis in the current reauthorization bill is on increasing the educational qualifications of Head Start teachers. Table 7 examines the effect of increasing specific educational inputs holding per capita expenditures constant.

The first panel indicates that higher pupil-teacher ratios are associated with lower scores

on tests of reading and vocabulary, and with a higher probability of grade repetition. The second

panel examines the effect of pupil-staff ratios, since most Head Start classrooms have both

teachers and aides. Lower pupil-staff ratios are also estimated to increase test scores, although

results are weaker, suggesting that teachers have a more positive effect on children than aides.

This is some evidence in support of the view that teacher qualifications matter.

However, the fraction of qualified teachers has no significant impact on outcomes, nor

does the teacher's salary, as shown in Panels 3 and 4.[20] It is possible that the null effect for

teacher qualifications reflects the fact that the fraction with qualifications is already very

high–by 1999, over 90% of teachers had qualifications, so the perception that the average Head

Start teacher is unqualified is unfounded. We also found little evidence that higher teacher pay

mattered, as shown in the next panel.

In the fifth panel we look at whether the average teacher is reported to have a B.A. or

higher degree. Again, the interaction of this variable with the "ever Head Start" indicator is not

statistically significant. Thus, given that most teachers already have a qualification such as a

child development certificate, there seems to be little measurable advantage to increasing their

education to the BA level.

We next examine several characteristics of Head Start directors. Recall that director

quality is a possible confounding factor, if directors who are able to obtain more funds, also run

programs that are better in other respects. However, we find no evidence that director education,

experience, or salary are positively related to outcomes among Head Start children. We have

---

[20] Similarly, Angrist and Guryan (2003) find that teacher certification in elementary schools raises teacher salaries, but has no impact on their quality as measured by SAT scores.

estimated similar models examining the effects of having a director who is full-time rather than part-time, and examining the average teacher's experience, and also find little effect on Head Start children's outcomes.

On the whole then, these results provide some support for the argument that increasing the share of expenditures devoted to educational programming would benefit Head Start children. But they suggest that educational dollars would be better spent reducing pupil-teacher ratios rather than on increasing teacher qualifications and/or salaries. Moreover, the estimates indicate that expenditures on health programming (broadly defined to include nutrition and programming for the disabled) is also beneficial, so that reallocating funds from health programs to education might prove counter-productive.

This finding supports the Head Start advocate's contention that it is difficult to educate children who are sick, hungry, or who have undiagnosed learning disabilities, and is consistent with a large body of evidence that poor health in childhood has negative impacts on educational attainment via mechanisms such as reduced schooling attendance (see Grossman and Kaestner (1997) for a survey). However, these results also raise the question of why Head Start health spending is important given that health and nutrition services are widely available through other programs such as WIC (the Supplemental Feeding Program for Women, Infants, and Children), Food Stamps, EPDST (the Early and Periodic Diagnostic Screening and Treatment Program which is available to Medicaid-eligible children), and state and federal programs for the disabled. It is possible that Head Start acts as a "case coordinator" and helps insure that children who need these other programs actually receive the benefits to which they are entitled. Alternatively, the combination of specialized services with preschool education may be

especially helpful.

VI. Discussion and Conclusions

There are several important limitations of our work. First, although the NLSY sample is large by the standards of Head Start research, it is small relative to the number of children in Head Start, and omits some categories of children (such as Hispanics whose mothers were not in the United States in 1978) entirely. Moreover, participation in Head Start is based on maternal reports, and is likely to be measured with some error.   Similarly, the administrative data available to us was incomplete and subject to measurement error. These limitations suggest that further research with better data is warranted.

Second, we are asking whether spending more on children who are currently in Head Start would improve their outcomes. It is important to keep in mind that Head Start has never been fully funded, and spending more on the children who get into the program rather than simply expanding the program might not be the best use of additional funds, particularly in light of evidence that the program as currently run does have lasting beneficial effects on children (c.f. Garces et al, 2002).

Third, there are at least two reasons to suspect that our results are lower bounds on the effects of Head Start spending. As discussed above, the Department of Health and Human Services is required to consider the neediness of the area when allocating funds, so it is possible that higher spending programs serve systematically needier children. Second, there is a good deal of measurement error in these data, as discussed above, which is likely to bias estimated effects towards zero.

Still, we believe that our results have implications beyond the current debate over the re-authorization of Head Start. In 1995, 31 percent of America's three year-olds, 61 percent of four year-olds, and 90 percent of five year-olds received some form of center-based care or attended kindergarten (National Center for Education Statistics, 1996). Head Start has served as a model for state preschools targeted to low-income children in states such as California (U.S. General Accounting Office, 1995), and also for new (voluntary) universal preschool programs in Georgia and New York. The Children's Defense Fund (1999) reports that as of the 1998-99 school year, 724,610 children were participating in state-funded enriched preschool programs. Hence, the number of children in state-funded early education initiatives is roughly equal to the 800,000 participants in Head Start and it is more important than ever before to determine what works, and what doesn't in early childhood intervention programs.

This study represents a first attempt to "get inside the black box" of Head Start program design to answer specific questions about the effectiveness of the program. Our results provide evidence consistent with both sides of the debate over the future of Head Start. On the one hand, we find evidence that higher spending programs are more effective. In particular, Head Start children in higher spending programs have larger gains on reading scores, and a lower probability of repeating grades. It has been argued that learning to read is the most important academic milestone for elementary school children, since all other learning rests on this foundation (National Research Council, 1998).

Although we need to be cautious about linear extrapolations, our estimates imply that funding Head Start programs at a per child level similar to the Perry Preschool Program (an increase in expenditures of roughly 50 percent), would essentially eliminate the gap in reading

achievement scores between the average Head Start child and other children. Moreover, the effects of spending on test scores are largest for children in poor counties, and among black children, suggesting that disadvantaged children would benefit disproportionately from increases in spending.

On the other hand, we also find evidence consistent with the contention that if the goal of Head Start is to improve child outcomes, then Head Start dollars should be more targeted towards services for children (health and education) and less targeted at other services (such as programs for parents and social programs).   Our findings suggest that a reallocation of funds along these lines would increase test scores and reduce behavior problems and grade repetition. Of course, it could be argued that providing services for parents and improving the home environment is one of the mandates of Head Start. This study is silent on the question of whether higher Head Start spending benefits parents and other household members.

Our examination of specific educational inputs suggests that if the share of Head Start dollars spent on education was increased, it would be more productive to reduce pupil/teacher ratios than to further increase teacher qualifications and salaries. While this finding cannot be the last word on this very controversial subject[21], it does suggest that legislators should be cautious about abrogating local control of budgets by requiring that Head Start funds be spent in a particular way (e.g. on hiring teachers with B.A. degrees) in the absence of any evidence that

---

[21] Hanushek (2002) summarizes a large literature finding small or inconsistent effects of pupil-teacher ratios in elementary schools, while Krueger (2002) argues that the better studies tend to find positive effects of class size reductions. Studies of quality in child care centers have shown that pupil/teacher ratios are only weakly related to indexes of quality such as the Early Childhood Environment Rating Scale (ECRS), but it is not known how predictive these scales are of latter achievement (Blau and Currie, forthcoming).

these constraints will have the intended effect on child outcomes.

References

The Abell Foundation. 2000 "The Untapped Potential of Baltimore City Public Preschools" (The Abell Foundation: Baltimore Maryland), June.

Angrist, Joshua and Jonathan Guryan. "Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements" NBER Working Paper #9545, 2003.

Barnett, Steven. 1995. "Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes" The Future of Children. 5:3, 25-50.

Blau, David.M. and Currie, Janet. forthcoming. "Who's Minding the Kids? Day Care, Preschool and Afterschool Care", in Eric Hanushek and Finis Welch (eds.) Handbook of Education Economics (Amsterdam: North Holland), forthcoming.

Children's Defense Fund. Seeds of Success: State Pre-K Initiatives 1998-1999, Washington D.C.: Children's Defense Fund, 1999.

Currie, Janet. "Early Childhood Intervention Programs: What Do We Know?", Journal of Economic Perspectives,15 #2, Spring 2001, 213-238.

Currie, Janet and Duncan Thomas. "Early Test Scores, Socioeconomic Status, School Quality and Future Outcomes," Research in Labor Economics, v 20, 2001, 103-132.

Currie, Janet and Duncan Thomas. "School Quality and the Longer-Term Effects of Head Start", Journal of Human Resources, fall 2000, v35 #4, 755-774.

Currie, Janet and Duncan Thomas. "Does Head Start Help Hispanic Children?", Journal of Public Economics, 74 #2, November, 1999, 235-262.

Currie, Janet and Duncan Thomas. "Does Head Start Make A Difference?," The American Economic Review, June 1995, 85 #3, 341-364.

Department of Health and Human Services, Administration for Children and Families, Head Start Bureau. "Head Start PCCost: The Computerized Version of the Grant Application Package SF424A Version 4.01 Y2K", Sept. 1999.

Department of Health and Human Services, Administration for Children and Families. "Policies and Procedures for Selection, Initial Funding, and Refunding of Head Start Grantees, and for Selection of Replacement Grantees", in Head Start Performance Standards, Section 1302. Http://www.acf.hhs.gov/programs/hsb/performance/1302.htm, October 3, 2001.

Department of Health and Human Services, Administration for Children and Families. "FY 2002 Head Start Funding Opportunities for Special Initiatives",

http://www.headstartinfo.org/publications/im02/im02_06a.htm, April 9, 2002.

Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. "Catalog of Federal Domestic Assistance", Section 93.6, http://aspe.os.dhhs.gov/cfda/p93600.htm, June 2002.

Department of Health and Human Services, Administration for Children and Families. "Funding Opportunities", http://www.acf.hhs.gov/programs/hsb/grant/fundingopportunities/fundopport.htm, Sept. 8, 2003.

Fewell, R.R., & K.G. Scott. 1997. "The Cost of Implementing the Intervention" in Helping low birth weight, premature babies: The Infant Health and Development Program. R.T. Goss, D. Spiker, & C.W. Haynes, eds. Stanford, CA: Stanford University Press, 479-502.

Garces, Eliana, Duncan Thomas, and Janet Currie. "Longer Term Effects of Head Start", The American Economic Review, v92 #4, Sept. 2002, 999-1012.

Grossman, Michael and Robert Kaestner. "The Effects of Education on Health" in The Social Benefits of Education, edited by Jere Behrman and Nevzer Stacey (Ann Arbor: University of Michigan Press) 1997, 69-123.

Krueger, Alan. 2002. "Economic Considerations and Class Size", NBER Working Paper #8875, April.

Hanushek, Eric. "Publicly Provided Education" in Alan Auerbach and Martin Feldstein (eds.) Handbook of Public Economics, (Amsterdam: North Holland, 2002), 2054-2141.

Heckman, James J., JingJing Hsse and Yona Rubinstein. 2000. "The GED is a Mixed Signal: The Effect of Cognitive and Non-Cognitive Skills on Human Capital and Labor Market Outcomes", University of Chicago xerox.

Lee, Valerie and Susanna Loeb. 1995. "Where do Head Start Attendees End Up? One Reason Why Preschool Effects Fade Out" Educational Evaluation and Policy Analysis. 17:1, 62-82.

Karoly, Lynn et al. 1998. Investing in our Children: What we Know and Don't Know About the Costs and Benefits of Early Childhood Interventions. Santa Monica: RAND.

National Center for Education Statistics. The Condition of Education, 1996, Washington D.C.: National Center for Education Statistics, 1996.

NICHD Early Child Care Research Network and Duncan, G.J. (2002) Modeling the Impacts of Child Care Quality on Children's Preschool Cognitive Development presented at Society for Research on Child Development, Minneapolis, April 2001.

Resnick, Gary and Nicholas Zill. undated. "Is Head Start Providing High-Quality Educational Services? Unpacking Classroom Processes", xerox, Westat, Inc.

National Research Council. "Preventing Reading Difficulties in Young Children". Catherine E. Snow, Susan Burns, and Peg Griffin (eds.) (National Academy Press: Washington D.C.) 1998.

U.S. Administration for Children Youth and Families. 1999. "Head Start Fact Sheet, 1998". Washington D.C.: Head Start Bureau.

U.S. Bureau of Labor Statistics. 2000. NLSY79 Child and Young Adults User's Guide 2000 http://www.bls.gov/nls/y79cyaguide/nlsy79cusg.htm.

U.S. Dept. of Health and Human Services. 2001. Head Start Bureau Program Performance Standards and Other Regulations. http://www2.acf.dhhs.gov/programs/hsb/performance/index.htm.

U.S. General Accounting Office. 1995. Early Childhood Centers: Services to Prepare Children for School Often Limited. Washington D.C.: Government Printing Office, GAP/HEHS-95-21, March.

Zill, Nicholas, Gar Resnick, and Ruth Hubbell McKey. undated. "What Children Know and Can Do at the End of Head Start and What it Tells us About the Program's Performance". xerox, Westat Inc.

Table 1: Comparison Agency Characteristics in Administrative Data Set and in NLSY Sample

	1	2	3	4	5	6	7	8	9
Variable	Agencies w Funding	Agencies w/o Fund.	Percap>= $4,000	Percap <$4000	NLSY All	NLSY Poor county	NLSY White	NLSY Black	NLSY Hispanic
Enrollment	487	473	481	491	704	860	560	830	877
	[522]	[510]	[494]	[542]	[669]	[841]	[481]	[717]	[885]
Per Capita Funding	3934		4978	3109	3538	3384	3599	3465	3495
	[1276]		[1094]	[653]	[1260]	[1163]	[1231]	[1238]	[1351]
Per Cap. Funding + In Kind	5251		6575	4204	4793	4619	4894	4748	4611
	[1649]		[1417]	[894]	[1981]	[1543]	[2097]	[1999]	[1609]
% spent - education	0.409	0.407	0.390	0.424	0.416	0.429	0.406	0.419	0.438
	[0.093]	[0.092]	[0.086]	[0.096]	[0.098]	[0.102]	[0.095]	[.104]	[0.096]
% spent - total health related	0.118	0.114	0.119	0.118	0.109	0.107	0.112	0.105	0.108
	[.046]	0.043	[0.046]	[0.046]	[0.040]	[0.039]	[0.040]	[0.044]	[0.036]
% spent - other	0.473	0.479	0.491	0.458	0.475	0.464	0.482	0.477	0.453
	[.091]	0.088	[0.083]	[0.093]	[0.094]	[0.100]	[0.090]	[0.103]	[0.089]
Fraction white, non Hisp.	0.510	0.587	0.413	0.587	0.375	0.291	0.493	0.230	0.296
	[0.385]	[0.350]	[0.398]	[0.358]	[0.324]	[0.302]	[0.340]	[0.241]	[0.278]
Fraction black	0.281	0.291	0.267	0.288	0.414	0.441	0.344	0.661	0.223
	[0.326]	[.0329]	[0.314]	[0.332]	[0.329]	[0.357]	[0.299]	[0.281]	[0.248]
Fraction Hispanic	0.113	0.114	0.120	0.109	0.185	0.244	0.131	0.095	0.446
	[0.209]	[0.209]	[0.203]	[0.212]	[0.250]	[0.306]	[0.185]	[0.164]	[0.311]
Pupil/classroom staff ratio	9.778	9.733	9.143	10.125	10.970	10.711	11.010	10.935	10.923
	[2.690]	[2.890]	[2.306]	[2.818]	[3.145]	[3.07]	[3.027]	[3.156]	[3.40]
Pupil/teacher ratio	22.732	22.394	21.097	23.639	24.527	23.573	25.505	23.251	24.013
	[9.784]	[9.282]	[8.498]	[10.319]	[8.623]	[8.691]	[9.352]	[7.432]	[8.10]
Teacher Salary	16427.53	15180.71	17797.01	15850.33	18198.37	18168.85	18151.70	17708.48	19024.54
	[5687]	5550.25	[5212.59]	[5779.13]	[6773.39]	[6242.18]	[6625.13]	[6957.69]	[6777.65]
Fraction qualified teachers	0.848	0.865	0.868	0.837	0.857	0.861	0.866	0.840	0.862
	[0.186]	[0.183]	[0.173]	[0.193]	[.171]	[0.168]	[0.169]	[0.161]	[0.185]
Average teacher has BA+	0.222	0.318	0.224	0.217	0.280	0.241	0.309	0.266	0.230
	[0.416]	[0.466]	[0.417]	[0.412]	[0.449]	[0.428]	[0.462]	[0.442]	[0.421]
Director has BA+	0.644	0.809	0.627	0.672	0.793	0.763	0.777	0.842	0.759
	[0.479]	[0.393]	[0.484]	[0.470]	[0.405]	[0.425]	[0.416]	[0.365]	[0.428]
Director's Years Experience	10.306	10.090	10.439	10.092	9.449	10.137	9.141	9.815	9.669
	[8.545]	[8.489]	[8.641]	[8.386]	[7.587]	[7.485]	[7.646]	[7.861]	[6.985]
Director's Salary	36876.55	37558.69	35183.74	39577.62	41698.00	41463.72	40477.64	42805.18	43007.73
	[11772.99]	[12529.12]	[11559.39]	[11606.81]	[12585.26]	[13618.33]	[11635.05]	[13042.72]	[13777.50]
# Observations	14532	4734	6419	8113	4468	1984	2238	1324	906

Notes: Standard deviations in brackets.

Table 2: Means in NLSY Sample

	1 Ever Head St.	2 PIAT Math	3 PIAT RR	4 PIAT-RC	5 PPVT	6 Behavior Problems	7 Repeat Grade	8 Mother AFQT
Panel 1: Sample Means								
1. All	0.227	51.03	58.26	54.27	37.69	57.58	0.123	36.38
		[23.85]	[24.66]	[23.97]	[28.68]	[25.17]	[.329]	[27.09]
2. Counties with >11% Families in Poverty	0.265	46.88	54.28	50.39	29.96	59.24	0.154	28.87
		[23.26]	[24.74]	[23.73]	[26.63]	[24.46]	[.361]	[24.72]
3. Whites	0.142	59.25	64.17	61.27	51.09	55.51	0.086	50.97
		[22.54]	[23.37]	[22.56]	[26.70]	[25.83]	[0.281]	[25.80]
4. Blacks	0.388	41.33	51.76	46.09	22.01	60.52	0.182	19.89
		[22.22]	[24.50]	[22.75]	[22.70]	[24.43]	[.386]	[17.64]
5. Hispanics	0.204	45.28	53.42	50	29.79	58.32	0.127	24.43
		[22.49]	[24.52]	[24.10]	[26.28]	[23.95]	[0.334]	[21.34]
Panel 2: Mean Differences Head Start and Other Children								
1. All		-9.058	-8.729	-10.315	-13.868	3.696	0.122	-13.325
		[0.854]	[0.885]	[0.919]	[1.151]	[0.906]	[0.015]	[0.946]
2. Counties with >11% Families in Poverty		-7.288	-8.399	-9.322	-9.768	2.996	0.129	-11.294
		[1.182]	[1.258]	[1.291]	[1.515]	[1.248]	[0.022]	[1.233]
3. Whites		-6.657	-5.541	-5.854	-7.486	3.349	0.101	-7.447
		[1.392]	[1.446]	[1.586]	[1.925]	[1.591]	[0.023]	[1.555]
4. Blacks		-1.973	-5.867	-5.776	-5.205	3.35	0.091	-5.223
		[1.254]	[1.387]	[1.350]	[1.428]	[1.387]	[0.026]	[0.985]
5. Hispanics		-7.793	-7.255	-9.45	-8.531	-0.839	0.122	-5.077
		[1.874]	[2.050]	[2.114]	[2.461]	[1.989]	[0.033]	[1.752]
Panel 3: Mean Differences Head Start and Other Children Adjusted for Observables								
1. All		-0.75	-2.603	-3.737	-2.082	2.429	0.074	
		[0.809]	[0.851]	[0.868]	[1.027]	[0.953]	[0.016]	
2. Counties with >11% Families in Poverty		-0.92	-3.502	-4.676	-2.227	2.401	0.099	
		[1.147]	[1.220]	[1.231]	[1.380]	[1.290]	[0.024]	
3. Whites		-1.584	-1.213	-2.705	-0.85	3.85	0.046	
		[1.387]	[1.427]	[1.557]	[1.968]	[1.680]	[0.025]	
4. Blacks		1.884	-2.383	-2.705	-1.503	3.304	0.071	
		[1.235]	[1.334]	[1.265]	[1.338]	[1.438]	[0.027]	
5. Hispanics		-5.863	-5.387	-7.118	-3.999	-0.303	0.134	
		[1.844]	[1.964]	[2.021]	[2.342]	[2.037]	[0.035]	

Standard deviations in brackets in Panel 1. Standard errors in brackets in Panels 2 and 3.

Table 3: Effect of Head Start on Outcomes

	1	2	3	4	5	6
					Behavior	Repeated
	PIAT-Math	PIAT-RR	PIAT-RC	PPVT	Problems	Grade
Ever Head Start*per	0.397	1.749	1.428	1.677	0.137	0
capita spending	[0.569]	[0.598]	[0.631]	[0.738]	[0.677]	[0.012]
Per capita spending	-0.214	-0.302	-0.95	-0.621	0.127	0.018
	[0.394]	[0.414]	[0.419]	[0.496]	[0.466]	[0.007]
Ever Head Start?	-2.184	-8.936	-8.72	-7.789	1.929	0.074
	[2.214]	[2.326]	[2.370]	[2.714]	[2.627]	[0.044]
Mother's AFQT	0.246	0.245	0.274	0.315	0.073	-0.001
	[0.018]	[0.019]	[0.019]	[0.022]	[0.021]	[0.000]
Permanent Income	0.081	0.072	0.049	0.095	-0.19	-0.001
	[0.018]	[0.019]	[0.019]	[0.023]	[0.021]	[0.000]
Mother drop out	-0.829	-11.346	-8.399	3.883	-16.488	0.208
	[10.299]	[10.816]	[10.271]	[13.112]	[12.192]	[0.184]
Mother high school	3.235	-6.406	-2.547	7.449	-22.288	0.111
	[10.313]	[10.831]	[10.287]	[13.126]	[12.209]	[0.184]
Mother Some College	3.751	-4.302	-2.162	8.68	-21.711	0.092
	[10.324]	[10.842]	[10.298]	[13.136]	[12.221]	[0.184]
Mother College	6.423	-3.798	-2.547	11.357	-26.199	0.091
	[10.368]	[10.889]	[10.348]	[13.197]	[12.272]	[0.185]
Mother Black	-7.5	-0.619	-1.738	-10.957	0.748	-0.009
	[1.047]	[1.101]	[1.104]	[1.306]	[1.225]	[0.020]
Mother Hispanic	-3.871	-0.129	0.021	-5.593	0.216	-0.047
	[1.123]	[1.182]	[1.184]	[1.401]	[1.315]	[0.021]
Child Male	-0.804	-5.164	-3.699	-0.251	-1.506	0.029
	[0.624]	[0.656]	[0.657]	[0.779]	[0.733]	[0.012]
Child First Born	2.766	5.83	6.073	6.803	0.577	-0.01
	[0.695]	[0.731]	[0.727]	[0.858]	[0.817]	[0.013]
Child > 5 sibs	-7.408	-6.046	-2.943	-7.64	-2.86	0.042
	[1.761]	[1.850]	[1.967]	[2.286]	[2.092]	[0.032]
# sibs mother in 1979	-0.289	-0.027	-0.299	-0.752	-0.666	-0.002
	[0.182]	[0.191]	[0.191]	[0.227]	[0.215]	[0.003]
County population, 1990	-0.055	-0.079	-0.019	-0.065	-0.007	0
100,000s	[0.027]	[0.028]	[0.029]	[0.034]	[0.032]	[0.001]
% population black, 1990	11.85	6.709	1.673	-13.13	-6.27	-0.138
	[5.320]	[5.589]	[5.590]	[6.561]	[6.213]	[0.100]
% population Hispanic, 1990	0.019	0.025	0.036	-0.124	-0.057	0
	[0.054]	[0.057]	[0.056]	[0.067]	[0.064]	[0.001]
% births to teen moms, 1988	-0.246	-0.154	0.114	-0.115	0.272	0.002
	[0.148]	[0.155]	[0.156]	[0.185]	[0.174]	[0.003]
Median family income, 1989	0.021	-0.038	-0.077	0.081	0.408	0.005
	[0.115]	[0.120]	[0.121]	[0.144]	[0.134]	[0.002]
% families below poverty, 1989	-0.085	-0.198	-0.409	-0.029	0.291	0.005
	[0.180]	[0.189]	[0.189]	[0.223]	[0.211]	[0.003]
% votes for Democrat president, 1	-0.027	-0.005	0.063	0.11	-0.005	0.002
	[0.068]	[0.071]	[0.071]	[0.084]	[0.079]	[0.001]
% pop < 18 years, 1990	-0.059	-0.073	0.092	-0.067	0.051	0.001
	[0.221]	[0.233]	[0.234]	[0.279]	[0.258]	[0.004]
% pop > 64 years, 1990	-0.046	-0.002	0.108	0.116	0.052	-0.002
	[0.155]	[0.163]	[0.164]	[0.195]	[0.181]	[0.003]
Last year HS, 1988-89	2.882	-0.092	-0.757	2.057	0.527	0.082
	[1.175]	[1.237]	[1.232]	[1.359]	[1.384]	[0.023]
1990-91	1.507	-1.256	-1.497	0.795	0.147	0.078

	[1.201]	[1.264]	[1.259]	[1.388]	[1.410]	[0.023]
1992-93	3.467	1.153	3.72	3.139	-3.448	0.025
	[1.221]	[1.284]	[1.284]	[1.428]	[1.436]	[0.021]
1994-95	2.751	2.094	5.591	4.333	-5.767	-0.009
	[1.322]	[1.391]	[1.393]	[1.795]	[1.554]	[0.021]
1996-97	1.823	4.5	11.137	0.353	-11.089	0
	[1.490]	[1.568]	[1.618]	[2.146]	[1.752]	[0.000]
1998-99	4.826	7.2	16.213	4.22	-10.095	0
	[1.765]	[1.854]	[2.388]	[2.511]	[2.098]	[0.000]
Observations	4278	4272	3822	3372	4349	2972
R-squared	0.29	0.27	0.31	0.4	0.11	0.11

Note: Standard errors in brackets. Models also include state fixed effects.

Table 4: Effects of Head Start on Outcomes, Alternative Specifications and Samples

	1 PIAT-Math	2 PIAT-RR	3 PIAT-RC	4 PPVT	5 Behavior Problems	6 Repeated Grade
1. Counties with > 11% of Families in Poverty						
Ever Head Start*per	1.027	2.293	1.871	1.674	0.096	-0.003
capita spending	[0.908]	[0.965]	[1.021]	[1.136]	[1.025]	[0.020]
Per capita spending	-0.135	-0.279	-1	-1.221	-0.07	0.008
	[0.694]	[0.737]	[0.749]	[0.854]	[0.789]	[0.014]
Ever Head Start?	-4.475	-11.447	-10.899	-7.649	2.071	0.108
	[3.341]	[3.551]	[3.625]	[3.961]	[3.768]	[0.071]
AFQT	0.234	0.266	0.287	0.285	0	-0.001
	[0.029]	[0.031]	[0.030]	[0.035]	[0.032]	[0.001]
Observations	1907	1906	1721	1533	1925	1333
R-squared	0.27	0.27	0.3	0.37	0.16	0.13
2. White						
Ever Head Start*per	0.788	1.363	0.759	2.169	0.067	-0.041
capita spending	[0.961]	[0.987]	[1.109]	[1.415]	[1.160]	[0.019]
Per capita spending	-0.366	-0.036	-0.59	-0.361	0.6	0.013
	[0.562]	[0.578]	[0.598]	[0.766]	[0.681]	[0.009]
Ever Head Start?	-4.608	-6.502	-5.508	-8.771	3.522	0.197
	[3.970]	[4.076]	[4.477]	[5.534]	[4.796]	[0.075]
AFQT	0.249	0.237	0.272	0.31	0.084	-0.001
	[0.023]	[0.024]	[0.024]	[0.031]	[0.028]	[0.000]
Observations	2127	2124	1858	1632	2184	1475
R-squared	0.23	0.24	0.25	0.24	0.13	0.13
3. Black						
Ever Head Start*per	1.892	2.335	1.844	1.713	0.046	0.014
capita spending	[0.959]	[1.035]	[1.018]	[1.062]	[1.146]	[0.022]
Per capita spending	-1.166	-1.076	-1.579	-0.134	-0.294	0.018
	[0.814]	[0.879]	[0.832]	[0.882]	[0.961]	[0.017]
Ever Head Start?	-4.69	-10.51	-8.886	-7.132	3.156	0.024
	[3.559]	[3.842]	[3.645]	[3.734]	[4.229]	[0.080]
AFQT	0.238	0.29	0.347	0.395	0.027	-0.004
	[0.044]	[0.047]	[0.045]	[0.047]	[0.051]	[0.001]
Observations	1280	1280	1168	1038	1287	891
R-squared	0.2	0.25	0.3	0.3	0.12	0.16
4. Hispanic						
Ever Head Start*per	-1.361	0.596	0.715	1.972	-0.206	0.018
capita spending	[1.301]	[1.387]	[1.540]	[1.695]	[1.439]	[0.025]
Per capita spending	0.354	0.543	-0.836	-1.285	-0.555	0.022
	[0.864]	[0.921]	[0.975]	[1.114]	[0.965]	[0.016]
Ever Head Start?	-0.992	-7.554	-9.544	-10.6	0.46	0.069
	[5.011]	[5.340]	[5.646]	[6.183]	[5.543]	[0.097]
AFQT	0.266	0.266	0.231	0.273	0.094	0.001
	[0.047]	[0.050]	[0.051]	[0.059]	[0.052]	[0.001]
Observations	871	868	796	702	878	606
R-squared	0.25	0.28	0.29	0.33	0.17	0.16
5. Alternative Per Capita Spending Measure that Includes all Cash and In-Kind						
Ever Head Start*per	0.198	1.369	0.93	1.148	-0.045	0.003
capita spending	[0.432]	[0.454]	[0.476]	[0.565]	[0.512]	[0.009]
Per capita spending	-0.13	-0.086	-0.143	-0.06	0.169	0.008
	[0.201]	[0.211]	[0.206]	[0.262]	[0.237]	[0.003]

Ever Head Start?	-1.709	-9.257	-8.108	-7.348	2.647	0.061
	[2.252]	[2.365]	[2.400]	[2.787]	[2.664]	[0.044]
AFQT	0.246	0.245	0.274	0.315	0.073	-0.001
	[0.018]	[0.019]	[0.019]	[0.022]	[0.021]	[0.000]
Observations	4278	4272	3822	3372	4349	2972
R-squared	0.29	0.27	0.31	0.4	0.11	0.11

Notes: Standard errors in brackets. All models are specified similarly to those in Table 3.

Table 5: Effects of Spending on the Selection of Head Start Mothers

Panel A: Dependent variable=AFQT of mother.

	[1]	[2]	[3]	[4]	[5]
Subgroup:	**All**	**Poor**	**White**	**Black**	**Hispanic**
Ever Head Start*per	0.673	-0.558	2.479	1.878	-0.719
capita spending	[0.598]	[0.983]	[1.126]	[0.761]	[1.177]
Per capita spending	0.048	0.877	1.182	-1.382	-0.78
	[0.415]	[0.746]	[0.666]	[0.641]	[0.782]
Ever Head Start?	-8.308	-9.059	-18.174	-11.412	0.608
	[2.330]	[3.615]	[4.635]	[2.831]	[4.547]
Observations	4427	1961	2220	1310	897
R-squared	0.37	0.21	0.13	0.18	0.29

Panel B: Dependent variable=Child Low Birth Weight

Ever Head Start*per	-0.012	-0.009	-0.002	-0.025	-0.001
capita spending	[0.008]	[0.013]	[0.012]	[0.016]	[0.017]
Per capita spending	0.006	0	0.001	0.017	0.006
	[0.005]	[0.010]	[0.007]	[0.013]	[0.012]
Ever Head Start?	0.047	0.029	-0.015	0.097	-0.017
	[0.029]	[0.048]	[0.049]	[0.058]	[0.064]
Observations	4244	1813	2097	1205	827
R-squared	0.02	0.04	0.04	0.06	0.11

Notes: Standard errors in brackets. The models also included all of the independent variables shown in Table 3.

Table 6: Effects of Different Budget Components on Outcomes

	1	2	3	4	5	6
					Behavior	Repeated
All	PIAT-Math	PIAT-RR	PIAT-RC	PPVT	Problems	Grade
Ever Head Start *	0.44	1.824	1.454	1.736	0.039	0
Per capita spending	[0.572]	[0.600]	[0.633]	[0.741]	[0.679]	[0.006]
Per capita spending	-0.224	-0.292	-0.917	-0.643	0.135	-0.002
	[0.396]	[0.416]	[0.421]	[0.498]	[0.468]	[0.004]
Ever Head Start *	26.744	47.501	34.487	21.121	-35.164	-0.2
Share Health	[18.858]	[19.782]	[20.608]	[24.369]	[22.210]	[0.194]
Share Health	-12.682	-13.568	0.005	-9.799	10.826	0.185
	[10.165]	[10.674]	[10.901]	[13.011]	[11.897]	[0.105]
Ever Head Start *	-7.014	-5.247	-3.473	-11.866	-16.191	-0.212
Share Education	[7.935]	[8.332]	[8.507]	[9.825]	[9.303]	[0.081]
Share Education	-1.282	3.699	3.448	-6.395	-7.093	-0.004
	[4.507]	[4.735]	[4.733]	[5.686]	[5.277]	[0.046]
Ever Head Start	-2.28	-12.15	-11.071	-5.228	12.874	0.119
	[5.032]	[5.282]	[5.396]	[6.189]	[5.897]	[0.052]
Observations	4278	4272	3822	3372	4349	4365
R-squared	0.29	0.27	0.31	0.41	0.11	0.03
Counties with >11% Families in Poverty						
Ever Head Start *	1.143	2.431	1.822	1.786	0.01	-0.005
Per capita spending	[0.911]	[0.968]	[1.024]	[1.139]	[1.027]	[0.010]
Per capita spending	-0.242	-0.264	-0.894	-1.33	-0.122	0
	[0.698]	[0.741]	[0.753]	[0.859]	[0.793]	[0.007]
Ever Head Start *	40.142	63.61	26.394	21.559	-43.528	-0.391
Share Health	[25.620]	[27.225]	[28.746]	[32.542]	[29.047]	[0.275]
Share Health	-29.637	-19.744	14.478	-26.78	9.221	0.219
	[17.347]	[18.430]	[19.364]	[22.333]	[19.550]	[0.185]
Ever Head Start *	-5.016	5.63	-4.446	-2.391	-26.199	-0.295
Share Education	[10.511]	[11.171]	[11.329]	[12.530]	[11.791]	[0.111]
Share Education	-7.262	-1.733	0.666	-11.471	5.82	-0.029
	[6.897]	[7.336]	[7.223]	[8.240]	[7.753]	[0.073]
Ever Head Start	-7	-21.143	-11.725	-9.306	18.231	0.204
	[6.843]	[7.273]	[7.340]	[8.119]	[7.654]	[0.073]
Observations	1907	1906	1721	1533	1925	1945
R-squared	0.27	0.27	0.3	0.37	0.16	0.05
Counties with >11% Families in Poverty -- Share of Health and Education Together						
Ever Head Start *	1.048	2.334	1.864	1.729	0.034	-0.004
Per capita spending	[0.910]	[0.967]	[1.022]	[1.136]	[1.024]	[0.010]
Per capita spending	-0.206	-0.265	-0.996	-1.298	-0.115	-0.001
	[0.695]	[0.739]	[0.750]	[0.855]	[0.790]	[0.007]
Ever Head Start *	-1.857	10.157	-1.299	-1.042	-27.589	-0.294
Share on Ed&Health	[10.325]	[10.978]	[11.134]	[12.285]	[11.585]	[0.110]
Share on Ed&Health	-8.358	-2.513	1.442	-12.062	5.958	-0.014
	[6.838]	[7.276]	[7.170]	[8.196]	[7.676]	[0.073]
Ever Head Start	-3.543	-17.014	-10.183	-7.305	17.027	0.192
	[6.575]	[6.992]	[7.024]	[7.721]	[7.335]	[0.070]
Observations	1907	1906	1721	1533	1925	1945
R-squared	0.27	0.27	0.3	0.37	0.16	0.04

Notes: Standard errors in brackets. All models include the other independent variables in Table 3.

Table 7: Effects of Individual Educational Inputs

	1	2	3	4	5	6
					Behavior	Repeated
	PIAT-Math	PIAT-RR	PIAT-RC	PPVT	Problems	Grade
Pupil-Teacher Ratio						
Ever Head Start *	0.648	2.017	1.907	1.637	0.081	-0.005
Per Capita Expenditure	[0.591]	[0.623]	[0.656]	[0.770]	[0.703]	[0.012]
Per Capita Expenditure	-0.118	-0.399	-1.124	-0.421	-0.204	0.016
	[0.418]	[0.440]	[0.443]	[0.522]	[0.494]	[0.008]
Pupil-Teacher Ratio	-3.149	-9.604	-11.145	-8.249	-2.116	0.106
* Ever Head Start	[3.397]	[3.573]	[3.604]	[4.191]	[4.003]	[0.064]
Pupil-Teacher Ratio	0.003	-0.013	0.037	0.022	0.172	-0.001
	[0.095]	[0.100]	[0.100]	[0.118]	[0.111]	[0.002]
Observations	4204	4198	3766	3323	4277	2929
R-squared	0.29	0.27	0.31	0.41	0.11	0.11
Pupil-Classroom Staff Ratio						
Ever Head Start *	0.597	1.921	1.853	1.6	0.14	-0.005
Per Capita Expenditure	[0.594]	[0.625]	[0.659]	[0.774]	[0.706]	[0.012]
Per Capita Expenditure	-0.023	-0.379	-1.074	-0.41	-0.155	0.017
	[0.421]	[0.443]	[0.445]	[0.526]	[0.497]	[0.008]
Pupil-Classroom Staff Ratio	-2.538	-7.311	-9.966	-7.541	-4.479	0.11
* Ever Head Start	[3.844]	[4.040]	[4.094]	[4.778]	[4.520]	[0.072]
Pupil-Classroom Staff Ratio	-0.029	-0.203	-0.008	-0.003	0.577	-0.002
	[0.255]	[0.268]	[0.271]	[0.320]	[0.299]	[0.005]
Observations	4205	4199	3767	3324	4278	2930
R-squared	0.29	0.27	0.31	0.4	0.11	0.11
Fraction of Teachers with Qualifications						
Ever Head Start *	0.655	2.068	1.951	1.762	-0.127	-0.005
Per Capita Expenditure	[0.591]	[0.621]	[0.654]	[0.771]	[0.703]	[0.012]
Per Capita Expenditure	0.02	-0.302	-1.048	-0.444	-0.135	0.013
	[0.414]	[0.435]	[0.437]	[0.516]	[0.488]	[0.008]
Fraction Qualified Teachers	-3.331	-4.298	-7.654	-8.028	4.391	-0.004
* Ever Head Start	[4.681]	[4.924]	[4.912]	[5.869]	[5.518]	[0.090]
Fraction Qualified Teachers	0.387	3.079	1.122	2.026	-2.287	0.001
	[2.249]	[2.367]	[2.347]	[2.774]	[2.633]	[0.041]
Observations	4205	4199	3767	3323	4278	2930
R-squared	0.29	0.27	0.31	0.41	0.11	0.11
Teacher Salary ($1,000)						
Ever Head Start *	0.677	2.185	2.177	2.052	0.144	0
Per Capita Expenditure	[0.703]	[0.740]	[0.785]	[0.956]	[0.826]	[0.014]
Per Capita Expenditure	0.023	-0.094	-0.713	-0.372	-0.132	0.022
	[0.481]	[0.506]	[0.514]	[0.610]	[0.564]	[0.009]
Lead Teacher Salary	-0.159	-0.079	-0.143	-0.137	0.003	-0.002
* Ever Head Start	[0.128]	[0.134]	[0.140]	[0.164]	[0.151]	[0.003]
Lead Teacher Salary	0.194	0.067	0.081	-0.072	-0.151	0.002
	[0.068]	[0.071]	[0.072]	[0.087]	[0.078]	[0.001]
Observations	4026	4019	3613	3196	4092	2803
R-squared	0.29	0.27	0.31	0.41	0.11	0.12

Table 7, continued

	1	2	3	4	5	6
					Behavior	Repeated
	PIAT-Math	PIAT-RR	PIAT-RC	PPVT	Problems	Grade
Average Teacher BA or More Education						
Ever Head Start *	0.41	1.765	1.415	1.543	0.054	-0.001
Per Capita Expenditure	[0.576]	[0.606]	[0.637]	[0.747]	[0.684]	[0.012]
Per Capita Expenditure	-0.209	-0.3	-0.936	-0.594	0.133	0.018
	[0.394]	[0.414]	[0.419]	[0.496]	[0.467]	[0.007]
Teacher BA+	-1.084	-0.782	-0.505	1.86	1.743	0.011
* Ever Head Start	[1.658]	[1.740]	[1.800]	[2.084]	[1.949]	[0.032]
Teacher BA+	-1.89	-8.75	-8.502	-7.84	1.711	0.073
	[2.223]	[2.335]	[2.382]	[2.726]	[2.639]	[0.044]
Observations	4278	4272	3822	3372	4349	2972
R-squared	0.29	0.27	0.31	0.4	0.11	0.11
Director BA or More Education						
Ever Head Start *	0.35	1.705	1.335	1.677	0.251	0.001
Per Capita Expenditure	[0.571]	[0.599]	[0.633]	[0.740]	[0.679]	[0.012]
Per Capita Expenditure	-0.183	-0.283	-0.96	-0.629	0.051	0.017
	[0.396]	[0.415]	[0.420]	[0.497]	[0.467]	[0.007]
Director BA+	-1.959	-1.989	-4.814	-0.375	4.354	0.029
* Ever Head Start	[1.891]	[1.986]	[2.031]	[2.509]	[2.225]	[0.035]
Director BA+	0.955	0.563	-0.84	-0.64	-2.414	-0.013
	[0.983]	[1.032]	[1.027]	[1.301]	[1.155]	[0.018]
Observations	4278	4272	3822	3372	4349	2972
R-squared	0.29	0.27	0.31	0.4	0.11	0.11
Head Start Director's Salary						
Ever Head Start *	0.59	1.89	1.94	2.223	0.002	-0.004
Per Capita Expenditure	[0.633]	[0.664]	[0.706]	[0.836]	[0.751]	[0.013]
Per Capita Expenditure	-0.119	-0.163	-0.918	-0.52	0.034	0.015
	[0.424]	[0.444]	[0.448]	[0.531]	[0.498]	[0.008]
Head Start Director's Salary	0	0.034	-0.021	-0.158	-0.008	0
* Ever Head Start	[0.063]	[0.066]	[0.068]	[0.084]	[0.074]	[0.001]
Head Start Director's Salary	0.055	0.005	-0.002	0.053	-0.08	-0.001
	[0.035]	[0.037]	[0.037]	[0.045]	[0.041]	[0.001]
Observations	4148	4142	3712	3275	4219	2892
R-squared	0.29	0.27	0.31	0.41	0.11	0.11
Head Start Director's Years of Experience						
Ever Head Start *	0.623	2.005	1.828	1.675	-0.004	-0.005
Per Capita Expenditure	[0.588]	[0.619]	[0.654]	[0.767]	[0.699]	[0.012]
Per Capita Expenditure	0.039	-0.097	-0.908	-0.346	-0.206	0.013
	[0.411]	[0.433]	[0.436]	[0.514]	[0.485]	[0.008]
Director's Years of Experience	-0.122	-0.06	0.02	0.049	-0.115	-0.001
* Ever Head Start	[0.099]	[0.104]	[0.106]	[0.124]	[0.116]	[0.002]
Director's Years of Experience	0.038	0.08	0.085	-0.136	-0.022	0
	[0.051]	[0.053]	[0.053]	[0.064]	[0.059]	[0.001]
Observations	4203	4197	3761	3317	4274	2926
R-squared	0.29	0.27	0.31	0.41	0.11	0.11

Notes: Standard errors in brackets. All models include the other independent variables in Table 3.