Available online at www.sciencedirect.com
ScienceDirect
Economics of Education Review 26 (2007) 83–99
Economics of Education Review
www.elsevier.com/locate/econedurev

Getting inside the "Black Box" of Head Start quality: What matters and what doesn't

Janet Currie[a], Matthew Neidell[b,*]

[a]Department of Economics, Columbia University, 420 West 118th Street, New York, NY 10027, USA
[b]Department of Health Policy and Management, Columbia University, 600 West 168th Street, New York, NY 10032, USA

Received 27 January 2005; accepted 25 March 2005

Abstract

Critics of Head Start contend that many programs spend too much money on programs extraneous to children. On the other hand, Head Start advocates argue that the families of 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 and vocabulary scores where Head Start spending was higher. Holding per capita expenditures constant, children in programs that devoted higher shares of their budgets to child-specific expenditures have fewer behavior problems and are less likely to have been retained in grade.

JEL classification: I28; I38

Keywords: Efficiency; Expenditures

1. 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 the "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] Since 1965, federal funding has increased from $96 million to $5.3 billion by 2000.

There have been dozens of studies of Head Start and related preschool and early school enrichment programs (see Currie, 2001). These studies have

*Corresponding author. Tel.: +1 212 342 4522; fax: +1 212 305 3405.
E-mail address: mn2191@columbia.edu (M. Neidell).

[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

0272-7757/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2005.03.004

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, 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 child-centered and that in many centers funding has been diverted to programs for parents and community development (The Abell Foundation, 2000). Concern has also been expressed about low levels of teacher pay and qualifications. Current debates in the House and Senate also stress that the program should focus more exclusively on children, and specifically, on education.

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% of teachers have a bachelor's degree within 5 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.[2] Moreover, as Blau and Currie (2004) point out, the existing literature on child care quality is of limited help, given that it has not conclusively determined which measures of program quality have a causal effect on child outcomes.

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 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 4 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. This study represents the first attempt to link administrative data from Head Start to measures of future child outcomes.

We find that relative to other children who were in the same location at age four, Head Start children have higher reading and vocabulary scores where Head Start spending was higher. Holding per capita expenditures constant, children in programs that devoted higher shares to child-specific expenditures on education and health are less likely to be retained in grade and have fewer behavior problems. However, we find little evidence that mandating that a fixed budget be spent on particular inputs would improve the effectiveness of the program. In particular, teacher salaries and the fraction of

[2]The most recent federally sponsored study of Head Start is the Family and Child Experiences Survey (Zill, Resnick, & Hubbell-McKey, undated). Unfortunately this study took a short-term perspective, following children over the course of 1 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. Other studies of Head Start Quality examine whether programs measure up to a pre-conceived list of desirable characteristics (such as involving parents), not whether they have positive effects on children (cf. Shultz, Lopez, & Hochberg,1996).

teachers with BA degrees 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 some 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 2 provides some background information about Head Start, Section 3 describes the data, Section 4 discusses methods, results are presented in Section 5, and Section 6 offers some conclusions.

2. Background

Head Start is a federal–local matching grant program, so it is funded quite differently from most education programs in the US. Local grantees apply to the Department of Health and Human Services (DHHS) for funds to run their programs on a 3 year cycle. Head Start grantees must provide 20% 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, 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 (2% 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, 2002). 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. The total amount of money available under these initiatives is small relative to total Head Start spending but they could have significant impact on individual programs.

These considerations suggest that grantees and therefore individual centers 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 3 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. It is far from obvious that people who are good grant writers will necessarily be wonderful preschool directors. Nevertheless, we will test to see whether controlling for observable characteristics of program directors reduces the estimated effects of expenditures. 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 fact that the match may be in kind (e.g. the donation of space in a local church) may help to mitigate this potential source of bias. The inclusion of non-Head Start children as a control group also helps to control for this possibility, as long as the amenities affect both Head Start and non-Head Start children

similarly. This assumption is discussed further below.

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 (US DHHS, 2001, 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 is possible that children will be selected differently in high- and low-spending counties, a possibility that we investigate below.

3. 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 1988–2000.[3] They provide information about the educational qualifications and pay of staff. The PCCOST data spans the period 1990–2001 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 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.[4]

The NLSY data tracks the children of the roughly 6000 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.[5] For each child, 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).[6]

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. (There is no test in the NLSY that is administered both before and after a child would normally have attended Head Start.) We use all non-missing observations for each score, in order to preserve the maximum possible number of observations.

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 (cf. Heckman, Hsse, & Rubinstein, 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. Though no variable by itself is likely to be a perfect measure of children's

[3]A unit of observation in the administrative data is therefore defined by the grantee's number, the delegate agency's number, and the year. Since there can be multiple centers within a delegate agency, data is not available at the level of the individual center.

[4]To preserve sample size in the PCCOST data, which was not required by each program in every year, we went through the data casting backward and forward as many as 3 years to fill in missing observations. 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.

[5]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% of children are reported to have attended other preschools, so that some children attended both Head Start and other preschools.

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

outcomes, we find consistent results across this wide range of indicators.

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 4 years old. Second, we retained 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.[7] The main 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 children (Currie & Thomas, 1995).

We lose 403 children for whom information on Head Start attendance is missing, 292 children with poor matches to the administrative data, 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 4468 children.[8]

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 whether the 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.[9] Head Start centers are required to give breakdowns of the "on the balance sheet" portion of their budgets into 10 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, p. 27).

We include information about the race and ethnic breakdown of children served by the programs.[10] Information about some specific educational inputs that are often examined in the literature on school quality are included: pupil/teacher ratios, teacher salaries, the fraction of teachers with qualifications, the average education of Head Start teachers, and

[7]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.

[8]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 one program. Since geographical data is not available at a finer level than the county in the NLSY, 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, 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 3 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.

[9]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% of on the balance sheet funds on administration (US Administration for children youth and families, 1999).

[10]The Hispanic sample in the NLSY includes only children born to mothers already in the country in 1978. Hence, it is not representative of the experience of more recent arrivals or of immigrant children.

Table 1
Comparison agency characteristics in administrative data set and in NLSY sample

Variable	1 Agencies w funding	2 Agencies w/o fund.	3 Percap> = $4000	4 Percap <$4000	5 NLSY all	6 NLSY poor county
Enrollment	487	473	481	491	704	860
	[522]	[510]	[494]	[542]	[669]	[841]
Per capita funding	3934		4978	3109	3538	3384
	[1276]		[1094]	[653]	[1260]	[1163]
Per cap. funding + in kind	5251		6575	4204	4793	4619
	[1649]		[1417]	[894]	[1981]	[1543]
% spent—education	0.409	0.407	0.390	0.424	0.416	0.429
	[0.093]	[0.092]	[0.086]	[0.096]	[0.098]	[0.102]
% spent—total health related	0.118	0.114	0.119	0.118	0.109	0.107
	[.046]	0.043	[0.046]	[0.046]	[0.040]	[0.039]
% spent—other	0.473	0.479	0.491	0.458	0.475	0.464
	[.091]	0.088	[0.083]	[0.093]	[0.094]	[0.100]
Fraction White, non-Hisp.	0.510	0.587	0.413	0.587	0.375	0.291
	[0.385]	[0.350]	[0.398]	[0.358]	[0.324]	[0.302]
Fraction Black	0.281	0.291	0.267	0.288	0.414	0.441
	[0.326]	[.0329]	[0.314]	[0.332]	[0.329]	[0.357]
Fraction Hispanic	0.113	0.114	0.120	0.109	0.185	0.244
	[0.209]	[0.209]	[0.203]	[0.212]	[0.250]	[0.306]
Pupil/classroom staff ratio	9.778	9.733	9.143	10.125	10.970	10.711
	[2.690]	[2.890]	[2.306]	[2.818]	[3.145]	[3.07]
Pupil/teacher ratio	22.732	22.394	21.097	23.639	24.527	23.573
	[9.784]	[9.282]	[8.498]	[10.319]	[8.623]	[8.691]
Teacher salary	16,427.53	15,180.71	17,797.01	15,850.33	18,198.37	18,168.85
	[5687]	[5550.25]	[5212.59]	[5779.13]	[6773.39]	[6242.18]
Fraction qualified teachers	0.848	0.865	0.868	0.837	0.857	0.861
	[0.186]	[0.183]	[0.173]	[0.193]	[.171]	[0.168]
Average teacher has BA +	0.222	0.318	0.224	0.217	0.280	0.241
	[0.416]	[0.466]	[0.417]	[0.412]	[0.449]	[0.428]
Director has BA +	0.644	0.809	0.627	0.672	0.793	0.763
	[0.479]	[0.393]	[0.484]	[0.470]	[0.405]	[0.425]
Director's years experience	10.306	10.090	10.439	10.092	9.449	10.137
	[8.545]	[8.489]	[8.641]	[8.386]	[7.587]	[7.485]
Director's salary	36876.55	37,558.69	35,183.74	39,577.62	41,698.00	41,463.72
	[11,772.99]	[12,529.12]	[11,559.39]	[11,606.81]	[12,585.26]	[13,618.33]
# Observations	14532	4734	6419	8113	4468	1984

Notes: Standard deviations in brackets.

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.

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.[11]

The last column of Table 1 shows means for the NLSY children in poor counties. 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% in their county, the median percent poor in our sample. Programs in poor counties tend to be much larger, and are slightly worse funded on average than those serving other children but are only slightly worse funded: average total funding per child is $4619 in poor counties compared to $4793 overall. This finding is 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, except that programs in poor counties place greater emphasis on educational expenditures.

Table 1 indicates that, despite concerns about teacher qualifications, over the whole sample period 85% 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. By 1997, we found that 92% of teachers had qualifications (up from 80% in 1991/92) and 34% of programs reported that their average teacher had a BA or higher (up from 22% in 1991/92). Although teachers receive what seems to be very low pay, averaging $16,428 per year, this salary implies an hourly wage rate of $17.50 per hour[12], which is comparable to the average hourly wage of $17.93 ($1998) that workers with BA degrees received in 2001 (authors' computation from the May Current Population Survey). The average salary received by Head Start program directors is $36,877 which is similar to the salary of a typical worker with a BA.[13]

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 8 months a year over 2 years. Since 20% of the children participated only for 1 year, the figures imply that the cost per child was approximately $7000 per year, so that Head Start costs about 71% 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 5 years.[14]

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. 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 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 dated. The NLSY children also have

[11]The pupil/teacher ratios in Table 1 are higher than the mandated maximum class size of 15–20 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 6 using the interaction of turnover rates with Head Start as our "input" and did not find any statistically significant effect.

[12]The data indicate that on average, Head Start teachers work 5 h per day (including preparation time), for 193 days during the year.

[13]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. 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.

[14]Fewell and Scott (1997) 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.

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]
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]
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]	

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

scores on the behavior problems index that are higher than the national norm, indicating a greater incidence of behavior problems. Children in poor counties tend to have lower test scores and more behavior problems and grade repetition than the full NLSY sample. Finally, 12.3% of children had ever repeated a grade by the time of the 2000 survey.

The second panel of Table 2 shows that Head Start children have significantly worse outcomes than other children. 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 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 Panels 2 and 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.

4. Methods

To examine the effect of Head Start spending on child outcomes, we estimate a reduced form education production function that includes per capita expenditures in place of specific Head Start inputs. We estimate least squares regression models of the following form:

$$Y_{ic} = a + a_1\,\mathrm{Percap}_c + a_2\,\mathrm{HeadStart}_{ic} + a_3\,\mathrm{Percap}_c * \mathrm{HeadStart}_{ic} + a_4\,X_{ic} + a_5 Z_c + a_6 S + a_7\,\mathrm{Cohort}_{ic} + e_{ic}, \quad (1)$$

where Y_{ic} is the outcome of interest (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; *HeadStart* is an indicator equal to one if the child attended Head Start and zero otherwise; X is a vector of mother and child characteristics 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.[15]

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 and with child outcomes. For example, 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 (academic years 1988, 1990, 1992, 1994, 1996, and 1998), which allows outcomes to vary for children of different cohorts. Finally, e is an idiosyncratic error term.

Numerous studies have stressed concerns about omitted variables that may bias estimates of the effects of school resources on child outcomes (see, e.g. Hanushek, 1986). For example, even conditional on the included individual-, county-, and state-level variables controlled for in our analysis, counties with higher spending on Head Start could have unobserved characteristics that are associated with higher test scores for all children, such as generous community health programs in addition to Head Start. To overcome this concern, we include all children, regardless of whether or not they attend Head Start, and interact per capita expenditures with Head Start status. The main effect of per capita spending (*Percap*) captures the overall effect of being in a high spending county, for both Head Start and non-Head Start children. To the extent that counties that spend more on Head Start also have unobserved characteristics associated with higher test scores for all children, the effects will be captured by *Percap*.

The main effect of Head Start is used as a proxy for the unobserved characteristics of children that are associated with attending Head Start. Table 2 suggests the coefficient on HeadStart, a_2, is likely to be negative because children who attend Head Start are negatively selected relative to other children. Our main focus is on a_3, the coefficient on the interaction between *Percap* and *HeadStart*. 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.

There are two assumptions necessary to identify a_3. The first is that the effect of local amenities is the same for Head Start and non-Head Start children. Given that the two groups differ, as seen in Table 2, it is certainly plausible they may be differentially affected by local amenities. To assess the importance of this concern, we conduct all of our analyses separately for children who were in counties with high poverty rates. 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 show, our results are 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.

The second assumption necessary to identify a_3 is that children are selected into Head Start similarly in high and low spending counties. That is, expenditures on Head Start do not affect a family's decision to enroll their child in Head Start (or the program's way of picking children). This assumption is questionable if, for example, higher spending Head Start programs attract more able children, or are able to include both the neediest children and

[15]Several previous studies compare siblings in the NLSY in order to identify the effect of Head Start on outcomes (cf. Currie & Thomas, 1995, 2000; Garces et al., 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.

children who are higher up in the ability distribution. In this case, 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 examine the importance of this potential problem. First, we regress *HeadStart* on all of the covariates in (1), and ask whether per capita expenditures on Head Start affect enrollment probabilities. Second, we examine the way spending affects the selection of Head Start children in terms of their mother's AFQT score, which is one of the most powerful predictors of the children's test scores. Clearly, spending can have no causal impact on AFQT since in most cases it was measured before the child was born, so any "effect" of spending on maternal AFQT would represent a change in the way that children were selected into the program. These regression models are of the same form as (1) except that AFQT is the dependent variable, and we omit maternal AFQT, permanent income, and maternal education from the set of regressors. 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 g. 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 were less likely to select low birth weight children, then this could have an effect on measured outcomes. The results of these exercises suggest that our results are not driven by differences in selection between high and low spending counties, as shown below.

Next, we ask if the type of spending matters by estimating models that control not only for per capita expenditures, but also for the fraction of the budget spent directly on children (e.g. on education, health care, nutrition, and services for the disabled) as opposed to services for parents, social events, and costs such as administration, transportation, and rent. The model we estimate takes the following form:

$$\begin{aligned} Y_{ic} = {} & a + a_1\,\mathrm{Percap}_c + a_2\,\mathrm{HeadStart}_{ic} \\ & + a_3\,\mathrm{Percap}_c * \mathrm{HeadStart}_{ic} + a_4\,\mathrm{ShareChild}_c \\ & + a_5\,\mathrm{ShareChild}_c * \mathrm{HeadStart}_{ic} + a_6 X_{ic} \\ & + a_7 Z_c + a_8 S + a_9\,\mathrm{Cohort}_{ic} + v_{ic}. \end{aligned} \tag{2}$$

Additionally, 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:

$$\begin{aligned} Y_{ic} = {} & a + a_1\,\mathrm{Percap}_c + a_2\,\mathrm{HeadStart}_{ic} \\ & + a_3\,\mathrm{Percap}_c * \mathrm{HeadStart}_{ic} + a_4\,\mathrm{Input}_c \\ & + a_5 \mathrm{Input}_c * \mathrm{HeadStart}_{ic} + a_6 X_{ic} + a_7 Z_c \\ & + a_8 S + a_9\,\mathrm{Cohort}_{ic} + v_{ic}, \end{aligned} \tag{3}$$

where *Input* is one of the educational inputs and the other variables are defined as above.

Models (2) and (3) differ from conventional production functions which examine the effect of changing one input holding all others constant. To change one input with all others constant would require changes in the budget. Instead, we want to ask what would happen if we took a fixed budget, and required that part of it be spent in a particular way. Although this deviates from the production function approach, we believe this exercise corresponds more closely to the policies actually under discussion.

5. Results

5.1. 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 (in real 1998 dollars) as reflected in Table 1. 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 children if per capita spending on the

Table 3
Effect of Head Start on outcomes

	1 PIAT-math	2 PIAT-RR	3 PIAT-RC	4 PPVT	5 Behavior problems	6 Repeated grade
Ever Head Start*per capita spending	0.397 [0.569]	1.749 [0.598]***	1.428 [0.631]**	1.677 [0.738]**	0.137 [0.677]	0 [0.012]
Per capita spending	−0.214 [0.394]	−0.302 [0.414]	−0.95 [0.419]**	−0.621 [0.496]	0.127 [0.466]	0.018 [0.007]**
Ever Head Start?	−2.184 [2.214]	−8.936 [2.326]***	−8.72 [2.370]***	−7.789 [2.714]***	1.929 [2.627]	0.074 [0.044]*
Mother's AFQT	0.246 [0.018]***	0.245 [0.019]***	0.274 [0.019]***	0.315 [0.022]***	0.073 [0.021]***	−0.001 [0.000]***
Permanent income	0.081 [0.018]***	0.072 [0.019]***	0.049 [0.019]***	0.095 [0.023]***	−0.19 [0.021]	−0.001 [0.000]*
Mother drop out	−0.829 [10.299]	−11.346 [10.816]	−8.399 [10.271]	3.883 [13.112]	−16.488 [12.192]	0.208 [0.184]
Mother high school	3.235 [10.313]	−6.406 [10.831]	−2.547 [10.287]	7.449 [13.126]	−22.288 [12.209]*	0.111 [0.184]
Mother some college	3.751 [10.324]	−4.302 [10.842]	−2.162 [10.298]	8.68 [13.136]	−21.711 [12.221]*	0.092 [0.184]
Mother college	6.423 [10.368]	−3.798 [10.889]	−2.547 [10.348]	11.357 [13.197]	−26.199 [12.272]*	0.091 [0.185]
Mother Black	−7.5 [1.047]***	−0.619 [1.101]	−1.738 [1.104]	−10.957 [1.306]***	0.748 [1.225]	−0.009 [0.020]
Mother Hispanic	−3.871 [1.123]***	−0.129 [1.182]	0.021 [1.184]	−5.593 [1.401]***	0.216 [1.315]	−0.047 [0.021]**
Child male	−0.804 [0.624]	−5.164 [0.656]***	−3.699 [0.657]***	−0.251 [0.779]	−1.506 [0.733]**	0.029 [0.012]*
Child first born	2.766 [0.695]***	5.83 [0.731]***	6.073 [0.727]***	6.803 [0.858]***	0.577 [0.817]	−0.01 [0.013]
Child > 5 sibs	−7.408 [1.761]***	−6.046 [1.850]***	−2.943 [1.967]	−7.64 [2.286]***	−2.86 [2.092]	0.042 [0.032]
# Sibs mother in 1979	−0.289 [0.182]	−0.027 [0.191]	−0.299 [0.191]	−0.752 [0.227]***	−0.666 [0.215]*	−0.002 [0.003]
% Population black, 1990	11.85 [5.320]**	6.709 [5.589]	1.673 [5.590]	−13.13 [6.561]**	−6.27 [6.213]	−0.138 [0.100]
% Population Hispanic, 1990	0.019 [0.054]	0.025 [0.057]	0.036 [0.056]	−0.124 [0.067]*	−0.057 [0.064]	0 [0.001]
% Births to teen moms, 1988	−0.246 [0.148]*	−0.154 [0.155]	0.114 [0.156]	−0.115 [0.185]	0.272 [0.174]	0.002 [0.003]
Median family income, 1989	0.021 [0.115]	−0.038 [0.120]	−0.077 [0.121]	0.081 [0.144]	0.408 [0.134]***	0.005 [0.002]**
% families below poverty, 1989	−0.085 [0.180]	−0.198 [0.189]	−0.409 [0.189]**	−0.029 [0.223]	0.291 [0.211]	0.005 [0.003]
% votes for Democrat president, 1992	−0.027 [0.068]	−0.005 [0.071]	0.063 [0.071]	0.11 [0.084]	−0.005 [0.079]	0.002 [0.001]
Observations	4278	4272	3822	3372	4349	2972
R^2	0.29	0.27	0.31	0.4	0.11	0.11

Note: Standard errors in brackets. Statistical significance at 10% (*), 5% (**) and 1% (***) level. Models also include county population in 1990, % of population < 18 years, % > 64 years, and cohort and state fixed effects.

program were reduced to zero (i.e. if there was no program).

At the mean level of spending of $3500, 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 would reduce gaps in verbal scores by between 1.4 and 1.7 points, so that an increase in

Table 4
Effects of spending on the selection of Head Start mothers

Panel A: Dependent variable = Ever attended Head Start	[1] All	[2] Poor
Per capita spending	0.004	0.014
	[.007]	[.013]
Observations	4427	1961
R^2	0.166	0.17

Panel B: Dependent variable = AFQT of mother or low birth weight.	[1]	[2]	[3]	[4]
Dep. var.	AFQT	AFQT	LBW	LBW
Subgroup:	All	Poor	All	Poor
Ever Head Start*per	0.673	−0.558	−0.012	−0.009
capita spending	[0.598]	[0.983]	[0.008]	[0.013]
Per capita spending	0.048	0.877	0.006	0
	[0.415]	[0.746]	[0.005]	[0.010]
Ever Head Start?	−8.308	−9.059	0.047	0.029
	[2.330]***	[3.615]***	[0.029]	[0.048]
Observations	4427	1961	4244	1813
R^2	0.37	0.21	0.02	0.04

Note: Standard errors in brackets. Statistical significance at 10% (*), 5% (**) and 1% (***) level. The models in Panel A include the set of independent variables specified in Table 3. The models in Panel B exclude AFQT, permanent income and mother's education.

spending of $1600 (i.e. total spending of $5100 "on the balance sheet") would eliminate the gap, if the effects of spending were linear.

Of the variables in our models, AFQT is among the most important determinants of child outcomes. The point estimates in Table 3 imply that test scores rise by about a quarter point for every one point gain in maternal AFQT, while the probability of grade repetition falls. Given AFQT, mother's education has little effect. Permanent income is also a highly significant determinant of test scores, though not of behavior problems or grade repetition. Race and ethnicity show a similar pattern, as does being first born. Children in larger families have lower test scores, and males have lower scores on the PIAT reading assessments. Several of the county-level variables, such as the percent of population that is black and the percentage of families below poverty, have significant effects on some of the outcomes. Lastly, the cohort effects indicate that the youngest cohorts have the highest test scores and the lowest incidence of behavior problems and grade repetition, as one would expect given that the oldest children in the NLSY tend to be born to the most disadvantaged families.[16]

Table 4 addresses the concern that children may be selected differently in high and low-spending counties. The first panel shows that per capita spending has no significant effect on the probability that a child attends Head Start in either the full set of counties or in poor counties. The second panel of Table 4 presents estimates of the "effects" of spending on maternal AFQT and low birth weight (with maternal AFQT, permanent income, and maternal education omitted from the right-hand side "X" vector). The first two columns show that

[16]Alternative specifications were also investigated. With 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 hold within this sample of relatively homogeneous counties, as well as in the full sample. This result suggests that the assumption that local amenities affect Head Start and non-Head Start children similarly is reasonable. Estimation was also performed 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. The results are qualitatively similar for both measures, though the point estimates are smaller using total per capita spending. In fact, the relative magnitudes are consistent with the interpretation that, on average, only spending "on the balance sheet" affects child outcomes. Details are available from the authors.

the interaction between per capita spending and "Ever Head Start" is not statistically significant in the AFQT model either in the whole or in the poor county sample. The next two columns of Panel B show that there is no relationship between spending and whether the selected child was of low birth weight.

Although we cannot rule out the possibility that Head Start children are selected very differently in terms of unobservables in high- and low-spending programs, the fact that spending has no detectable effect on either the probability that individual children attended or on the observable characteristics of children selected into the program suggests that this is unlikely. Hence, 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.

5.2. Getting inside the Black Box

Thus far, the results suggest 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 raising teacher qualifications, and less money on other "extraneous" programs. Other observers feel strongly that the best way to help children is by also helping their parents and communities by offering a holistic set 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 5 shows estimates from models which control both for per capita expenditures and for the *share* of the budget spent directly on services for children (i.e. the aggregate of the education, health, nutrition, and disability portions of the budget).[17] The interaction between Head Start attendance and per capita expenditures, shown in the first row, confirms that children who attended programs with higher per capita spending have higher reading and vocabulary test scores. The interaction between Head Start attendance and the share of budget spent on education and health, the third row of coefficients, shows however, that conditional on spending, programs that devote a higher share of the budget to children have lower grade repetition. The coefficient on behavior problems is also significant at the 90% level of confidence. The second panel of Table 5 shows that these effects are even more pronounced in poor counties.

The administrative data allow further investigation of whether any specific educational inputs can be shown to be beneficial. For example, emphasis in the 2003 version of the reauthorization bill (discussed above) was on increasing the educational qualifications of Head Start teachers. Table 6 examines the effect of increasing specific educational inputs holding per capita expenditures constant, which corresponds to this type of policy.

The interactions between "ever Head Start" and per capita Head Start expenditures are robust to the inclusion of additional input measures (omitted from the table). However, we find little evidence that reallocating fixed spending to particular outputs would have a beneficial effect. The first row of the first panel shows that that higher pupil–teacher ratios have little impact holding spending constant. The second panel examines the effect of pupil–staff ratios, since most Head Start classrooms have both teachers and aides. Again, there is little evidence of an effect. The third panel shows that the fraction of teachers who are qualified has no significant impact on outcomes, nor does the teacher's salary, as shown in Panel 4.[18] 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.

In the fifth panel we look at whether having the average teacher is reported to have a BA or higher degree affects children's outcomes. 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

[17] 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.

[18] 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.

Table 5
Effect of spending large budget shares on children

All	1 PIAT-math	2 PIAT-RR	3 PIAT-RC	4 PPVT	5 Behavior problems	6 Repeated grade
Ever Head Start*Per capita spending	0.38	1.731	1.414	1.679	0.076	0
	[0.571]	[0.600]***	[0.633]**	[0.740]**	[0.679]	[0.006]
Per capita spending	−0.22	−0.286	−0.934	−0.649	0.107	−0.003
	[0.395]	[0.415]	[0.420]**	[0.497]	[0.467]	[0.004]
Ever Head Start*Share on ed&health	−4.857	−1.856	−0.840	−9.507	−17.019	−0.205
	[7.834]	[8.230]	[8.402]	[9.684]	[9.178]*	[0.080]**
Share on ed&health	−1.86	2.826	3.375	−6.5	−6.015	0.007
	[4.470]	[4.698]	[4.701]	[5.657]	[5.231]	[0.046]
Ever Head Start	0.446	−7.906	−8.238	−2.77	11.128	0.116
	[4.817]	[5.060]	[5.162]	[5.891]	[5.651]**	[0.049]**
Observations	4278	4272	3822	3372	4349	4365
R^2	0.29	0.27	0.31	0.41	0.11	0.03
Counties with >11% families in poverty						
Ever Head Start*Per capita spending	1.048	2.334	1.864	1.729	0.034	−0.004
	[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*Share on ed&health	−1.857	10.157	−1.299	−1.042	−27.589	−0.294
	[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^2	0.27	0.27	0.3	0.37	0.16	0.04

Note: Standard errors in brackets. Statistical significance at 10% (*), 5% (**) and 1% (***) level. All models include the independent variables in Table 3.

to be little measurable advantage to increasing their average education to the BA level (from the current equivalent of an Associates degree).

In the next three panels, we examine the effects of 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 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.

The main effects of these specific inputs are generally statistically insignificant, with a few exceptions. For example, the significant main effects in the panel dealing with teacher salary suggest that lead Head Start teachers are paid more in areas where children have higher PIAT-math scores and a lower incidence of behavior problems, while the panel dealing with BA qualifications suggests that the average Head Start teacher is less likely to have a BA or more in areas where children have high reading and vocabulary scores.

On the whole, these results provide some support for the argument that increasing the share of expenditures devoted directly to children (e.g. to educational programming) would benefit Head Start children. But they suggest that mandating that local grantees spend funds in a very specific way would likely have little impact in the absence of new funding for the program.

6. 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 it omits some categories of children (such as Hispanics whose mothers were not in the United

Table 6
Effects of individual educational inputs

	1 PIAT-math	2 PIAT-RR	3 PIAT-RC	4 PPVT	5 Behavior problems	6 Repeated grade
Panel 1: Pupil–teacher ratio						
Pupil–teacher ratio*ever Head Start	0.003	−0.013	0.037	0.022	0.172	−0.001
	[0.095]	[0.100]	[0.100]	[0.118]	[0.111]	[0.002]
Pupil–teacher ratio	−0.083	−0.091	−0.051	−0.011	−0.033	0.002
	[0.049]*	[0.051]*	[0.051]	[0.060]	[0.057]	[0.001]*
Observations	4204	4198	3766	3323	4277	2929
R^2	0.29	0.27	0.31	0.41	0.11	0.11
Panel 2: Pupil-classroom staff ratio						
Pupil–classroom staff ratio*ever Head Start	−0.029	−0.203	−0.008	−0.003	0.577	−0.002
	[0.255]	[0.268]	[0.271]	[0.320]	[0.299]*	[0.005]
Pupil–classroom staff ratio	−0.109	−0.212	−0.114	−0.079	−0.068	0.005
	[0.134]	[0.141]	[0.138]	[0.163]	[0.157]	[0.002]*
Observations	4205	4199	3767	3324	4278	2930
R^2	0.29	0.27	0.31	0.4	0.11	0.11
Panel 3: Fraction of teachers with qualifications						
Fraction qualified teachers*ever Head Start	−3.331	−4.298	−7.654	−8.028	4.391	−0.004
	[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^2	0.29	0.27	0.31	0.41	0.11	0.11
Panel 4: Teacher salary ($1,000)						
Lead teacher salary*ever Head Start	−0.159	−0.079	−0.143	−0.137	0.003	−0.002
	[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^2	0.29	0.27	0.31	0.41	0.11	0.12
Panel 5: Average teacher BA or more education						
Teacher BA + *ever Head Start	−1.084	−0.782	−0.505	1.86	1.743	0.011
	[1.658]	[1.740]	[1.800]	[2.084]	[1.949]	[0.032]
Teacher BA +	1.490	0.897	1.242	0.922	−0.457	0.006
	[0.818]*	[0.860]	[0.857]	[1.009]	[0.960]	[0.015]
Observations	4278	4272	3822	3372	4349	2972
R^2	0.29	0.27	0.31	0.4	0.11	0.11
Panel 6: Director BA or More Education						
Director BA + *ever Head Start	−1.959	−1.989	−4.814	−0.375	4.354	0.029
	[1.891]	[1.986]	[2.031]**	[2.509]	[2.225]**	[0.035]
Director BA +	0.955	0.563	−0.840	−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^2	0.29	0.27	0.31	0.4	0.11	0.11
Panel 7: Head Start director's salary						
Head Start director's salary*ever Head Start	0.000	0.034	−0.021	−0.158	−0.008	0.000
	[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	−0.080	−0.001
	[0.035]	[0.037]	[0.037]	[0.045]	[0.041]	[0.001]
Observations	4148	4142	3712	3275	4219	2892
R^2	0.29	0.27	0.31	0.41	0.11	0.11
Panel 8: Head Start director's years of experience						
Director's years of experience*ever Head Start	−0.122	−0.060	0.020	0.049	−0.115	−0.001
	[0.099]	[0.104]	[0.106]	[0.124]	[0.116]	[0.002]

Table 6 (*continued*)

	1 PIAT-math	2 PIAT-RR	3 PIAT-RC	4 PPVT	5 Behavior problems	6 Repeated grade
Director's years of experience	0.038	0.080	0.085	−0.136	−0.022	0.000
	[0.051]	[0.053]	[0.053]	[0.064]**	[0.059]	[0.001]
Observations	4203	4197	3761	3317	4274	2926
R^2	0.29	0.27	0.31	0.41	0.11	0.11

Note: Standard errors in brackets. Statistical significance at 10% (*), 5% (**) and 1% (***) level. All models include the independent variables in Table 3.

States in 1978). 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 (cf. 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 DHHS is required to consider the neediness of the area when allocating funds, so it is possible that higher spending programs serve systematically needier children. In addition, the measurement error in these data may bias estimated effects towards zero.

Notwithstanding, this study represents a first attempt to "get inside the black box" of Head Start program design to answer specific questions about its effectiveness. 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 and vocabulary scores. 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%), 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 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 community development). Our findings suggest that a reallocation of funds along these lines would reduce behavior problems and grade repetition. Of course, it could be argued that serving parents and communities is one of the core mandates of Head Start. This study is silent on the question of whether higher Head Start spending benefits parents and other community members, and thus on the larger question of whether Head Start should be judged primarily in terms of its effects on children.

Our examination of specific educational inputs found little evidence that targeting existing dollars towards specific inputs would be effective. While this finding cannot be the last word on this very controversial subject,[19] it does suggest that legislators should be cautious about abrogating local

[19]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 & Currie, 2004).

control of budgets by requiring that Head Start funds be spent in a particular way (e.g. on hiring more teachers with BA degrees rather than more teachers with Associates degrees) in the absence of any evidence that these constraints will have the intended effect on child outcomes.

Acknowledgments

We thank Latika Chaudary and Graciana Rucci for excellent research assistance, Craig Turner for giving access to the Head Start administrative data, and Pedro Carneiro, participants at the NBER Summer Institute, the Minneapolis Federal Reserve Board's October 2003 Conference on the Economics of Early Childhood Development, Princeton University, the University of Texas, the editor, and two anonymous referees for excellent comments. The Administration for Children and Youth and Families generously provided the administrative data used in the paper but is not responsible for our results or conclusions.

References

The Abell Foundation. (2000). *The untapped potential of Baltimore city public preschools*. Baltimore: The Abell Foundation.

Angrist, J., & Guryan, J. (2003). Does teacher testing raise teacher quality? Evidence from state certification requirements. *NBER Working Paper* #9545.

Blau, D. M., & Currie, J. (2004). Who's minding the kids? Day care, preschool and afterschool care. In E. Hanushek, & F. Welch (Eds.), *Handbook of education economics*. Amsterdam: North-Holland.

Currie, J. (2001). Early childhood intervention programs: What do we know? *Journal of Economic Perspectives*, *15*(2), 213–238.

Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *The American Economic Review*, *85*(3), 341–364.

Department of Health and Human Services (DHHS), Administration for Children and Families, Head Start Bureau. Head Start (1999). PCCost: The computerized version of the grant application package SF424A version 4.01 Y2K.

Department of Health and Human Services (DHHS), Administration for Children and Families (2001). 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 (DHHS), Office of the Assistant Secretary for Planning and Evaluation (2002). Catalog of federal domestic assistance, section 93.6, http://aspe.os.dhhs.gov/cfda/p93600.htm, June 15, 2002.

Department of Health and Human Services (DHHS), Administration for Children and Families (2003). Funding opportunities. http://www.acf.hhs.gov/programs/hsb/grant/fundingopportunities/fundopport.htm, September 8, 2003.

Fewell, R. R., & Scott, K. G. (1997). The cost of implementing the intervention. In R. T. Goss, D. Spiker, & C. W. Haynes (Eds.), *Helping low birth weight, premature babies: The infant health and development program* (pp. 479–502). Stanford: Stanford University Press.

Garces, E., Thomas, D., & Currie, J. (2002). Longer term effects of Head Start. *The American Economic Review*, *92*(4), 999–1012.

Hanushek, E. (2002). Publicly provided education. In A. Auerbach, & M. Feldstein (Eds.), *Handbook of Public Economics* (pp. 2054–2141). Amsterdam: North Holland.

Hanushek, E. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, *24*(3), 1141–1177.

Heckman, J. J., Hsse J-J., & Rubinstein, Y. (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.

Krueger, A. (2002). *Economic considerations and class size*. NBER working paper #8875.

Karoly, L., Greenwood, P., Everingham, S., Hoube, J., Killburn, M. R., Rydell, C. P., 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.

Resnick, G., & Zill, N. (undated). Is Head Start providing high-quality educational services? Unpacking classroom processes. Westat, Inc. Xerox.

Schultz, T., Lopez E., Hochberg, M. (1996). *Early childhood reform in seven communities* (pp. 96-1320.). Washington, DC: US Dept. of Education Office of Educational Research and Improvement, ORAD.

National Research Council. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

US Administration for Children Youth and Families. (1999). *Head Start fact sheet*. Washington, DC: Head Start Bureau.

US Deptartment of Health and Human Services (US DHHS) (2001). *Head Start Bureau program performance standards and other regulations*. http://www2.acf.dhhs.gov/programs/hsb/performance/index.htm.

Zill, N., Resnick, G., Hubbell McKey, R. (undated). What children know and can do at the end of Head Start and what it tells us about the program's performance. Westat Inc. Xerox.