Title: Effects of Head Start hours on children's cognitive, pre-academic, and behavioral outcomes: An instrumental variable analysis

Authors and Affiliations:
Weilin Li, George Farkas, Greg J. Duncan, Deborah Lowe Vandell
University of California, Irvine

Margaret Burchinal
University of North Carolina at Chapel Hill

Abstract

Background / Context: Children from low-income families benefit remarkably from exposure to compensatory education that began with Head Start in 1965 and aimed to improve school readiness skills by design (Farran, 2007; Scarr & Weinberg, 1986). While empirical evidence has supported more instructional time in elementary and secondary schools for low-income students (Abdulkadiroglu, Angrist, Dynarski, Kane, & Pathak, 2011; Angrist, Dynarsky, Kane, Pathak, & Walters, 2010; Dobbie & Fryer, 2011; Hoxby, Muraka, & Kang, 2009; Patall, Cooper, & Allen, 2010), little is known that whether increasing quantity of Head Start could also benefit low-income children. Also largely unexamined is how Head Start quantity effects differ for different age groups.

Research Question: (1) Does the amount of daily exposure to Head Start impact cognitive, pre-academic, and social outcomes? (2) Does the impact vary by age?

Setting: The HSIS (National Head Start Impact Study) used a nationally representative sample of Head Start applicants to estimate the impacts of the program on children and their families (Advisory Committee on Head Start Research and Evaluation).

Population / Participants / Subjects: 4,442 applicants to 383 Head Start centers that participated in the HSIS.

Intervention / Program / Practice: In 2002, the HSIS randomly assigned 4,442 children who had applied to 383 Head Start centers to either treatment group that offered Head Start enrollment or control group that was not granted access to the Head Start centers they applied for during that academic year.

Research Design: Our goal is to estimate the effect of Head Start hours on child development. However, there could be selection bias, i.e. there could be other factors such as unobserved family and child characteristics that were both correlated with hours in Head Start centers and exert their own effect on child outcomes, leading to selection bias in our estimation of the effect of hours in Head Start centers.

We utilize the variance generated from random assignment and hours offered in Head Start centers using instrumental variable approach to address this problem by taking advantage of variance in center care hours that was generated by the HSIS random assignment. We conduct two-stage least square regression (henceforward IV) by using number of hours per day offered in Head Start centers as instrumental variables for the hours per day children spent in Head Start centers. We get the following regression equation:

$$Hours\ in\ HS = \gamma_1 Hours\ offered + \gamma_2 \cdot Center + \gamma_3 \cdot Contols + e\ .$$

This generates predicted values for hours in Head Start centers. In the second stage regression, we use predicted values of hours in Head Start centers as an independent variable and child outcomes as dependent variable:

$$Outcome = Predicted\ hours\ in\ HS \cdot \beta_1 + \beta_2 \cdot Center + \beta_3 \cdot Contols + u\ .$$

Because our independent variable of interest is not correlated with any family and child characteristics, we obtain the estimated effect of Head Start hours on child outcomes from this second stage regression. In other words, we take advantage of the amount of variation in Head

Start hours that was generated by hours offered in Head Start centers and estimate the effect of hours in Head Start centers on child outcomes. The analysis may not include control variables only if the random assignment was implemented perfectly so that the treatment and control groups were perfectly equivalent in all child and family characteristics. In this set of analysis, we include a set of control variables that we introduced before to adjust for any departures from random assignment and to get more precise estimates.

It could be possible that effects of Head Start hours could be non-linear. Hence we include a set of cubic regressions to detect non-linear effects (Marra & Radice, 2011). Specifically, the analysis include two stages:

1st stage regression is $Q = \gamma_1 TX \cdot Center + \gamma_2 Center + \gamma_3 Controls + e$;

2nd stage regression is $Y = \sum_i f_i\left(Q_i\right) + \beta_2 Center + \beta_3 Controls + \hat{e} + \varepsilon$,

where $f_i(\cdot)$ is the function of cubic regressions and $\hat{e}$ is the residual from the 1st stage regression including all unobservable information in the 1st stage. Residuals from the 1st stage regression $\hat{e}$ contain all unobservable information that can be used to obtain corrected parameter estimates of the focal variables of cubic function on hours in Head Start centers. In other words, $\hat{e}$ acts as a proxy variable in the 2nd stage regression.

Data Collection and Analysis:

Hours in Head Start centers. We create a variable of *hours per day in Head Start centers* from parent interview. In Spring 2003, parents reported their children' care settings and number of hours per week that their children spent in settings. For children whose child care settings were center care, we divide the original variable of "number of hours per week in settings" by 5 to create the variable of *hours per day in Head Start centers*; for children whose child care settings were not center care, we use 0 as their values for this variable.

Because about 10% of children in the HSIS sample entered Head Start centers after taking baseline assessment, we adjust Head Start hours by enrollment time by:

$$quant_{adjusted} = quant \cdot \frac{elapsed\ months\ between\ HS\ enrollment\ time\ \&\ spring\ assessment}{elapsed\ months\ between\ baseline\ assessment\ \&\ spring\ assessment}.$$

Hours offered by Head Start centers. We create a variable of *hours offered by Head Start centers* from center director interview. In Spring 2003, directors of Head Start centers reported the beginning time and ending time every day in their Head Start centers from Monday to Sunday. We calculate our variable of *hours offered in Head Start centers* from this report.

Child outcomes. We focus on four child outcomes that were assessed during spring 2003, roughly one academic year after the experiment began. The first outcome is the Peabody Picture Vocabulary Test (henceforward PPVT), third edition (Dunn & Dunn, 1997). The PPVT is an untimed test measuring receptive vocabulary. The examiner presents a series of four pictures to each child. The examiner states a word describing one of the pictures and asks the child to point to the picture that the word describes (reliability = .95). The second outcome is the Woodcock-Johnson III Tests of Achievement: Letter-Word Identification (henceforward WJ-letter word; Woodcock & Johnson, 1989; 1990). The WJ-letter word measures letter and word identification skills. The published median reliability of the WJ-letter word is 0.91 in the 5- to 19-age range. The third outcome is the Woodcock-Johnson III Tests of Achievement: Applied Problems (henceforward WJ-applied problems). This test measures the child's ability to analyze and solve practical math problems. To solve the problems that are read by the assessor to the child, the child must recognize the procedure to be followed and then count and/or perform simple calculations. The published median reliability is 0.92 in the 5-19 age range. And the last

outcome is parental reported child behavioral problems (Achenbach, 1987). Parents were asked to rate their children on items dealing with aggressive or defiant behavior, inattentive or hyperactive behavior and shy, withdrawn, or depressed behavior. For each item, the parent was asked to judge whether the behavioral description was "not true", "sometimes true", or "very true" of the child. This paper derives the total behavior problem scale from parent ratings containing 14 rating items, and the total scale score could range from zero (all items marked "not true") to 28 (all items marked "very true"). All of the outcomes used in this paper were normed to have mean 0 and standard deviation 1.

Covariates. We include baseline outcomes that were assessed at fall 2002 in the analysis. Other covariates in the analysis were also assessed at fall 2002, including a dummy variable for age cohort (1 = age 3 cohort), child gender (1 = male), child race / ethnicity (1 = black, 1 = Hispanic, 1 = White and other), whether Spanish was baseline testing language (1 = Spanish), child age at Spring assessment in weeks, maternal education (1 = less than high school, 1 = high school diploma or GED, 1 = beyond high school), whether the mother was married (1 = married mother), whether the mother was teenager (1 = teenager mother), whether biological parents lived together (1 = live together), whether the mother was a recent immigrant (1 = immigrant mother), and caregiver age.

Imputation and weights. We conduct single imputation on missing observations for academic achievement at Fall 2002 and for Head Start hours (offered and taken) between Fall 2002 and Spring 2003 using predictive mean matching method. Then we conduct our analysis using the spring 2003 final child weights (CHSPR2003WTCA).

Findings / Results: Tables 1 present weighted, post-imputation descriptive statistics by age cohorts. The columns show means and standard deviations for each cohort group. On average, children spent about 2 to 3 hours per day in Head Start centers.

[Insert Table 1 here]

Figure 1 presents histograms for Head Start hours by cohort. For both cohorts, time in Head Start centers ranged from 0 to 8 hours per day.

[Insert Figure 1 here]

Table 2 shows OLS and IV estimates for effects of Head Start hours on child outcomes. The first-stage regression includes hours offered in Head Start centers, center dummies, and the set of covariates. The F-statistics of hours offered in Head Start centers were above 600, which ensure sufficient variance of hours spent in Head Start that are generated by hours offered in Head Start centers.

[Insert Table 2 here]

PPVT. The first column of Table 2 shows estimates for effects Head Start hours on PPVT scores. It shows that IV generated higher standard errors than the OLS. The IV results indicate that for the full sample of both cohorts, an additional hour per day spent in Head Start centers increased PPVT scores by .062 SD (se = .011; $p < .001$). For the age-3 cohort, an additional hour per day spent in Head Start centers increased PPVT scores by .064 SD (se = .015; $p < .001$). This effect is similar to that for the age-4 cohort, whose magnitude is .045 SD (se = .017; $p < .001$). Joint test showed that differential effects between age-3 cohort and age-4 cohort are not statistically significant.

WJ-letter words. Column 2 of Table 2 shows estimates for effect of hours per day spent in Head Start centers on WJ-letter words. The IV results indicate that for the whole sample of both cohorts, an additional hour per day spent in Head Start centers increased WJ-letter words

scores by .086 SD (*se* = .014; $p < .001$). This effect is significantly larger for the age-3 cohort than for the age-4 cohort ($p < .05$).

WJ-applied problems. The third column displays effect estimates for hours per day spent in Head Start centers on WJ-applied problems. It shows that an additional hour per day spent in Head Start centers increases WJ-applied problems scores by .046 SD for the cohorts combined (*se* = .013; $p < .001$). The estimated effect is .042 SD for the age-3 cohort (*se* = .016; $p < .001$), and .057 SD for the age-4 cohort (*se* = .023; $p < .001$). Joint test showed that differential effects between age-3 cohort and age-4 cohort are statistically significant ($p < .05$).

Behavioral problems. The last column shows effect estimates of hours in center care on behavioral problems. The IV results show that an additional hour per day spent in Head Start centers decreases problem behaviors of children in age-3 cohort by .052 SD (*se* = .015; $p < .001$). Results do not show significant effects for behavioral problems of age-4 cohort. Joint test showed that differential effects between age-3 cohort and age-4 cohort are statistically significant ($p < .01$).

Non-linear effects. Table 3.1 and 3.2 display OLS and IV estimate for non-linear effects of Head Start hours on child outcomes. Significant results from IV approach were only found for outcome of WJ letter-word for age-3 and age-4 cohort. And it's surprising that, age-3 and age-4 cohorts that show similar patterns in the linear effects are different in non-linear effects.

[Insert Table 3.1 here]

[Insert Table 3.2 here]

Figure 2 provides predicted values from both linear and non-linear estimates. Left figures show comparison of predicted values from linear and non-linear estimates. And right figures show "zoom-in" details of predicted values from non-linear estimates with 95% confidence intervals.

Comparison between Figure 2a and 2b shows different pattern in Head Start hours' effects for age-3 cohort and for age-4 cohort. For age-3 cohort, WJ Letter-word went up quickly at the first 2 hours then went flat; while for age-4 cohort,t he score was flat at the first several hours then went up after 2 to 3 hours/day.

[Insert Figure 2 here]

Conclusions: Our results showed significant positive effects of hours in center care experienced on the child's cognitive, language, and academic outcomes. These results were consistent with Loeb et al. (2007) that showed significant positive effects of center care quantity on child outcomes. These results indicate that center-based education serves as a protective factor for the cognitive, language, and academic development of economically disadvantaged children. Therefore, findings support a strategy of increasing hours that children spend in center-based education.

In addition, we found significant effects of center care hours on maternal reported problem behaviors *only* for age-3 cohort. This, together with different patterns in non-linear effects for age-3 and age-4 cohorts on WJ Letter-word, suggests future exploration on heterogeneity treatment effects on children with varied level of sustained attention or temperament (e.g. shyness) that was suggested in Megan Gunnar's work about cortisol differences associated with child care hours and moderated by child age, temperament, and child care quality.

Appendix A. References

Achenbach, T. M., Edelbrock, C., & Howell, C. T. (1987). Empirically based assessment of the behavioral/emotional problems of 2-3-year old children. Journal of Abnormal Child Psychology, 15, 629-650.

Bereiter, C. (1986). Does direct instruction cause delinquency? *Early Childhood Research Quarterly*. 1,3:289-92.

Bereiter, C. & Englemann, S. (1966). Teaching the disadvantaged child in the preschool. Englewood Cliffs, NJ: Prentice-Hall.

Borenstein M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-analysis, Version 2. Englewood NJ: Biostat.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*(3).

Carnegie Corporation. (1994). A matter of time: Risk and opportunity in the out-of-school hours. New York, NY: Author.

Chattin-McNichols, John. (1992b). Montessori programs in public schools. ERIC Digest. Champaign, IL: ERIC Clearinghouse on Elementary and Early Childhood Education. ED 348 165.

Chubrick, R.E., and Kelley, M.F. Head Start expansion in the 1990's: A critique. Focus on Early Childhood (1994) 6:1-3.

Collins, R. (1990). Head Start research and evaluation: A blueprint for the future. Washington, DC: Administration for Children, Youth, and Families.

Cooper H., Allen A. B., Patall E. A., Dent A. L. (2010). Effects of full-day kindergarten on academic achievement and social development. Review of Educational Research, 80, 34–70.

Crandall, D. P., Bauchner, J. E., Loucks, S. F., & Schmidt, W. (1982). Models of the school improvement process: Factors Contributing to success. *4 Study of Dissemination EffortsSurporting School Improvement*. Andover, MA: The NETWORK, Inc.

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

Dobbie, Will, and Roland G. Fryer. (2012). Getting beneath the veil of effective schools: Evidence from New York City. NBER Working Paper No. 17632.

Dodge, D. T., Colker, L. J., & Heroman, C. (2002). The creative curriculum for preschool (4th ed.). Washington, DC: Teaching Strategies.

Dunn, L. M. & Dunn L. M. (1997). *Examiner's Manual for the Peabody Picture Vocabulary Test-Third Edition.* Circle Pines, MN: American Guidance Service, Inc..

Farran, D. C. (2007). Is education the way out of poverty? A reflection on the 40th anniversary of Head Start (with commentaries by James King and Bernard L. Charles), Monographs of the Center for Research on Children's Development & Learning, No. 3.

Fusaro, J, A. (1997). The Effect of Full-Day Kindergarten on Student Achievement: A Meta-Analysis. Child Study Journal, Vol. 27, No. 4.

Gatawa, B. S. M. (1990). *The Politics of the School Curriculum: An Introduction*. Harare: Jongwe Press.

Gersten, R. (1986) Response to "Consequences of three preschool curriculum models through age 15." *Early Childhood Research Quarterly*, 1,3:293-302.

Goffin, S. G. & Wilson, C. S. (2001). Curriculum models andearly childhood education: Appraising the relationship, 2nd, Upper Saddle, NJ:Merrill/Prentice Hall.
Gullo, D. F. (2000). The Long-Term Educational Effects of Half-Day vs. Full-Day Kindergarten. Early Childhood Development and Care, 160: 17–24.
Haskins, R. The efficacy of early childhood education. American Psychologist (1989) 44,2:274-82.
Hoxby, Caroline M., Sonali Murarka, and Jenny L. Kang. (2009). How New York City's charter schools affect achievement. Cambridge, MA: New York City Charter Schools Evaluation Project.
Illinois State Board of Education. (2009). Prekindergarten / preschool for all curriculum criteria.
Karnes, M.B., Schwedel, A.M., & Williams, M.B. (1983). A comparison of five approaches for educating young children from low-income homes. In *As the twig is bent, lasting effects of preschool programs*. Consortium for Longitudinal Studies, ed. Hillsdale, NJ: Erlbaum, pp. 133-70.
Kisker, E. E., Hofferth, S. L., Phillips, D. A., Farquhar, E. (1991). A profile of childcare settings: Early education and care in 1990. Princeton, NJ: Mathematica Policy Research, Inc., 1991.
Jacob, R. T., Creps, C. L., & Boulay, B. (2004). *Meta-analysis of research and evaluation studies in early childhood education.* Cambridge, MA: Abt Associates Inc..
Layzer, J. I., Goodson, B. D., Bernstein, L., & Price, C. (2001). *National evaluation of family support programs, volume A: The meta-analysis, final report.* Cambridge, MA: Abt Associates Inc.
Layzer, J. I., Layzer, C. J., Goodson, B. D. & Price, C. (2007). Evaluation of Child Care Subsidy Strategies: Findings from Project Upgrade in Miami-Dade County.
Lee, V. E., Burkham, D. T., Ready, D. D., Honigman, J., & Meisels, S. J. (2006). Full-day versus half-day kindergarten: In which programs do children learn more? American Journal of Education, 112, 163–208.
Lipsey, Farran, Hurley, Hofer, & Bilbrey, (2009). Presentation in the annual conference of the Society of Research in Educational Effectiveness.
Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. Economics of Education Review, 26, 52–66.
Lonigan, C. J., Clancy-Menchetti, J., Phillips, B. M., McDowell, K., & Farver, J. M. (2005). *Literacy express: A preschool curriculum.* Tallahassee, FL: Literacy Express.
McCartney, K., Burchinal, M., Clarke-Stewart, A., Bub, K. L., Owen, M. T., Belsky, J., & NICHD ECCRN. (2010). Testing a series of causal propositions relating time in child care to children's externalizing behavior. Developmental Psychology, 46, 1-17.
Miller, L. B. & Bizzell, R. P. (1983). The Louisville Experiment: A comparison of four programs. In *As the twig is bent, lasting effects of preschool programs*. Consortium for Longitudinal Studies, ed. Hillsdale, NJ: Erlbaum, pp. 171-200.
Miller, L.B., & Bizzell, R.P. (1984) Long-term effects of four preschool programs: Ninth and tenth grade results. *Child Development*. 55:1570-87.
NAEYC (2002). National Association for the Education of Young Children and the National Council of Teachers of Mathematics. www.naeyc.org/resources/position_statements/psmath.htm.

NCES Digest of Education Statisitcs (2010). Retrieved from NCES website, http://nces.ed.gov/programs/digest/.
NICHD-ECCRN. (2003). Does amount of time spent in child care predict socio-emotional adjustment during the transition to kindergarten? Child Development, 74(4), 976–1005.
Office of School Readiness, State Board of Education, North Carolina. (2008). North Carolina approved early childhood curricula. Retrieved from http://www.osr.nc.gov/_pdf/NCApprovedEarly ChildhoodCurricula.pdf
O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, 78(1), 33-84.
Patall, E. A., Cooper, H. & Allen, A. B. (2010). Extending the School Day or School Year: A Systematic Review of Research (1985 – 2009). Review of Educational Research. September 2010. Vol. 8, No. 3, pgs. 401-436.
Puma, M. S. Bell, R. Cook, & Heid, C. (2010). Head Start impact study: Final report. Westat. Report prepared for the U.S. Department of Health and Human Services.
Sarama, J. & Clements, D.H. (2004). Building blocks of early childhood mathematics. *Teaching Children Mathematics*, 9, p. 480–484.
Scarr, S. & Weinberg, R. (1986). The early childhood enterprise: Care and education of the young. *American Psychologist*, 41, pp. 1140–1146.
Schweinhart, L.J., Weikart, D.P., & Larner, M.B. (1986). Consequences of three preschool curriculum models through age 15. *Early Childhood Research Quarterly.* 1,1:15-46.
Shager, H. M., Schindler, H. S., Hart, C. M. D., Duncan, G. J., Magnuson, K. A., & Yoshikawa, H. (2010). Using Meta-Analysis to Explain Variation in Head Start Research Results: The Role of Research Design. The SREE Annual Research Conference, Washington, D.C., March 4.
Steven, J. R., & Taylor, A. M. (2009). Hierarchical dependence in meta-analysis. *Journal of educational and behavioral statistics*, 34, 46-73.
Stipek, D. J. (2006). No Child Left Behind comes to preschool. *Elementary School Journal*, 106, 455–465.
Stipek, D., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development,* 66, 209-223.
Walston, J. & West, J. (2004). Full-Day and Half-Day Kindergarten in the United States: Findings from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99. NCES 2004-078.
Woodhead, M. When psychology informs public policy: The case of early childhood intervention. American Psychologist (1988) 43,6:443-54.
Woodcock, R. W., & Johnson, M. B. (1990). *Tests of achievement, WJ–R. Examiner's manual.* Allen, TX: DLM Teaching Resources.

Appendix B. Tables and Figures

Table 1. Descriptive statistics, weighted, after imputation.

	Both cohorts			Age 3 cohort		Age 4 cohort	
	Mean	SD	Imputed counts	Mean	SD	Mean	SD
Child characteristics - Baseline (Fall 2002)							
Age 3 cohort	54.6%						
Gender - male	49.1%			48.0%		50.4%	
Race							
Black	30.4%			34.3%		25.6%	
Hispanic	36.2%			32.9%		40.2%	
White & Other	33.4%			32.8%		34.2%	
Spanish as baseline test language	23.5%			19.4%		28.3%	
Age at spring assessment in weeks	236.29	29.78		214.77	17.52	262.14	18.94
Family characteristics - Baseline (Fall 2002)							
Maternal education							
Less than high school	37.1%			34.2%		40.6%	
High school diploma / GED	29.2%			31.2%		26.8%	
Beyond high school	33.7%			34.6%		32.6%	
Married mother	45.3%			45.3%		45.3%	
Teenage mother	14.5%			13.7%		15.5%	
Parents lived together	50.8%			50.3%		51.3%	
Immigrant mother	18.4%			15.2%		22.2%	
Caregiver age	29.2	7.21		29.13	7.37	29.28	7.02
Academic achievement - Baseline (Fall 2002)							
PPVT	91.52	8.9	55	91.79	7.56	91.2	10.27
WJ Letter words	90.23	11.4	321	90.97	11.61	89.34	11.09
WJ Applied problems	93.07	13	359	93.28	13.51	92.81	12.34
Behavioral problems	6.05	3.63		6.05	3.58	6.05	3.68
Head Start hours (Fall 2002 - Spring 2003)							
Hours offered in Head Start centers	4.64	5.04	467	4.46	4.86	4.85	5.24
Hours spent in Head Start centers	1.83	2.07	98	2.00	2.17	1.63	1.93
Academic achievement – One year follow-up (Spring 2003)							
PPVT	92.09	9.53		92.32	8.44	91.81	10.68
WJ Letter words	89.69	13.02		90.05	12.66	89.26	13.44
WJ Applied problems	93.84	13.05		94.1	13.53	93.53	12.45
Behavioral problems	5.84	3.63		5.99	3.66	5.66	3.59
Sample size	3540			1693		1577	

Table 2. OLS and IV estimates of effects of Head Start hours on child outcomes.

			PPVT	WJ Letter-word	WJ Applied Problems	Behavioral Problems
Both cohorts	OLS	β	.028 ***	.062 ***	.035 ***	-.029 ***
		(*se*)	(.007)	(.009)	(.008)	(.008)
		Controls	*Included*	*Included*	*Included*	*Included*
		R^2	.59	.47	.45	.41
	2SLS	β	.062 ***	.086 ***	.046 ***	-.042 ***
		(*se*)	(.011)	(.014)	(.013)	(.013)
		Controls	*Included*	*Included*	*Included*	*Included*
		1st stage F	1794	1799	1796	1795
Age 3 cohort		β	.025 **	.067 ***	.036 ***	-.029 ***
		(*se*)	(.009)	(.011)	(.009)	(.010)
		Controls	*Included*	*Included*	*Included*	*Included*
		R^2	.56	.49	.51	.47
	2SLS	β	.064 ***	.097 ***	.042 ***	-.052 ***
		(*se*)	(.015)	(.017)	(.016)	(.015)
		Controls	*Included*	*Included*	*Included*	*Included*
		1st stage F	1018	1014	1005	1019
Age 4 cohort		β	.025 **	.056 ***	.030 +	-.033 *
		(*se*)	(.009)	(.012)	(.016)	(.013)
		Controls	*Included*	*Included*	*Included*	*Included*
		R^2	.73	.55	.54	.47
	2SLS	β	.045 ***	.069 ***	.057 ***	-.026
		(*se*)	(.017)	(.022)	(.023)	(.020)
		Controls	*Included*	*Included*	*Included*	*Included*
		1st stage F	689	687	679	683

Notes: + $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$. Control variables included in regressions are: child gender, race/ethnicity, baseline testing language, child age, maternal education level, whether the mother was married, whether the mother was teenager, whether parents lived together, whether the mother was recently immigrant, caregiver age, and baseline outcomes assessed at Fall, 2002. Spring 2003 final child weights were used in analyses. Higher scores in cognitive outcomes and academic achievement indicate better outcomes; higher scores in behavioral outcomes indicate more negative development. Clustered standard errors were calculated using pairs cluster bootstrap (Cameron, Gelbach, & Miller, 2008) at center level with 600 replicates.

Table 3.1. OLS estimates of non-linear effects of Head Start hours on child outcomes

			PPVT		WJ Letter-word		WJ Applied Problems	Behavioral Problems	
Both cohorts	Hr/d	β	.160	*	.065		.002	-.088	
		(*se*)	(.071)		(.083)		(.071)	(.088)	
	Hr/d ^{2}	β	-.066	+	-.013		.012	.009	
		(*se*)	(.035)		(.042)		(.036)	(.042)	
	Hr/d ^{3}	β	.008	+	.003		-.001	.001	
		(*se*)	(.004)		(.005)		(.004)	(.005)	
	Controls		*Included*		*Included*		*Included*	*Included*	
	R^2		.59		.47		.46	.41	
Age 3 cohort	Hr/d	β	.258	*	.349	***	-.057	-.237	*
		(*se*)	(.120)		(.108)		(.114)	(.103)	
	Hr/d ^{2}	β	-.117	*	-.140	**	.029	.073	
		(*se*)	(.057)		(.053)		(.055)	(.051)	
	Hr/d ^{3}	β	.014	*	.017	**	-.001	-.006	
		(*se*)	(.007)		(.006)		(.006)	(.006)	
	Controls		*Included*		*Included*		*Included*	*Included*	
	R^2		.56		.49		.51	.48	
Age 4 cohort	Hr/d	β	.006		-.281	*	.005	.006	
		(*se*)	(.085)		(.116)		(.117)	(.132)	
	Hr/d ^{2}	β	.017		.166	***	.019	-.014	
		(*se*)	(.044)		(.057)		(.058)	(.063)	
	Hr/d ^{3}	β	-.003		-.019	***	-.003	.001	
		(*se*)	(.005)		(.007)		(.007)	(.007)	
	Controls		*Included*		*Included*		*Included*	*Included*	
	R^2		.73		.56		.54	.47	

Notes: + $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$. Samples are restricted to less than 6 hours per day. Control variables included in regressions are: child gender, race/ethnicity, baseline testing language, child age, maternal education level, whether the mother was married, whether the mother was teenager, whether parents lived together, whether the mother was recently immigrant, caregiver age, and baseline outcomes assessed at Fall, 2002. Spring 2003 final child weights were used in analyses. Higher scores in cognitive outcomes and academic achievement indicate better outcomes; higher scores in behavioral outcomes indicate more negative development. Clustered standard errors were calculated using pairs cluster bootstrap (Cameron, Gelbach, & Miller, 2008) at center level with 600 replicates.

Table 3.2. IV estimates of non-linear effects of Head Start hours on child outcomes

			PPVT		WJ Letter-word		WJ Applied Problems	Behavioral Problems	
Both cohorts	Hr/d	β	.122	+	.032		-.011	-.079	
		(*se*)	(.068)		(.084)		(.070)	(.090)	
	Hr/d ^{2}	β	-.038		.012		.021	.003	
		(*se*)	(.034)		(.044)		(.036)	(.044)	
	Hr/d ^{3}	β	.005		.000		-.001	.001	
		(*se*)	(.004)		(.005)		(.004)	(.005)	
	Controls		*Included*		*Included*		*Included*	*Included*	
Age 3 cohort	Hr/d	β	.190		.311	***	-.072	-.209	+
		(*se*)	(.126)		(.110)		(.123)	(.110)	
	Hr/d ^{2}	β	-.076		-.117	*	.038	.057	
		(*se*)	(.061)		(.055)		(.061)	(.055)	
	Hr/d ^{3}	β	.010		.014	*	-.002	-.004	
		(*se*)	(.007)		(.006)		(.007)	(.006)	
	Controls		*Included*		*Included*		*Included*	*Included*	
Age 4 cohort	Hr/d	β	-.009		-.301	*	-.015	.003	
		(*se*)	(.085)		(.118)		(.122)	(.124)	
	Hr/d ^{2}	β	.030		.185	***	.037	-.012	
		(*se*)	(.044)		(.057)		(.062)	(.059)	
	Hr/d ^{3}	β	-.004		-.021	***	-.004	.001	
		(*se*)	(.005)		(.007)		(.007)	(.007)	
	Controls		*Included*		*Included*		*Included*	*Included*	

Notes: + $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$. Samples are restricted to less than 6 hours per day. Control variables included in regressions are: child gender, race/ethnicity, baseline testing language, child age, maternal education level, whether the mother was married, whether the mother was teenager, whether parents lived together, whether the mother was recently immigrant, caregiver age, and baseline outcomes assessed at Fall, 2002. Spring 2003 final child weights were used in analyses. Higher scores in cognitive outcomes and academic achievement indicate better outcomes; higher scores in behavioral outcomes indicate more negative development. Clustered standard errors were calculated using pairs cluster bootstrap (Cameron, Gelbach, & Miller, 2008) at center level with 600 replicates.

Figure 1. Histograms of hours spent in Head Start centers.

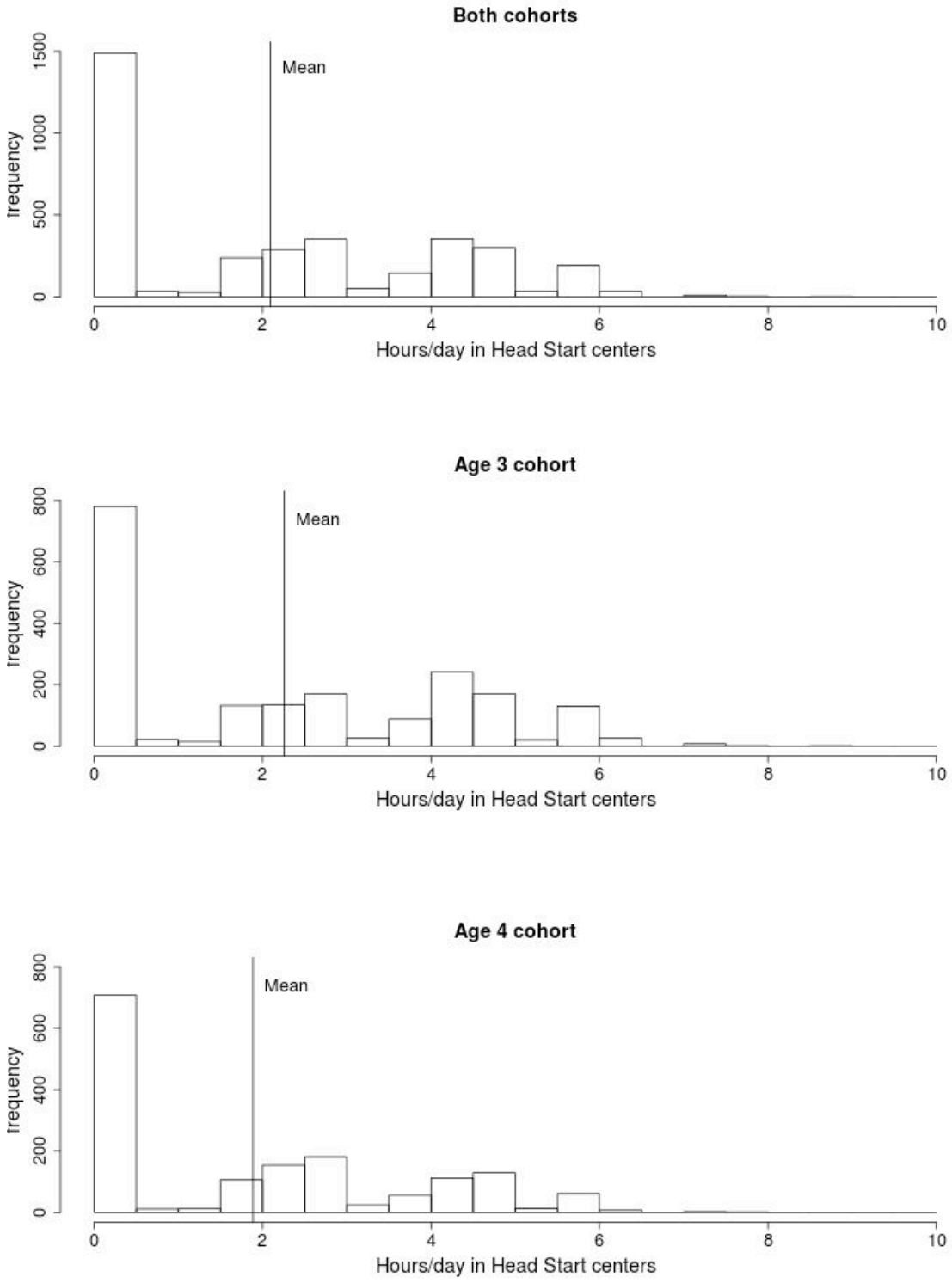

Figure 2a. Linear and non-linear effects of Head Start hours from Instrumental Variable approach: Age 3 cohort.

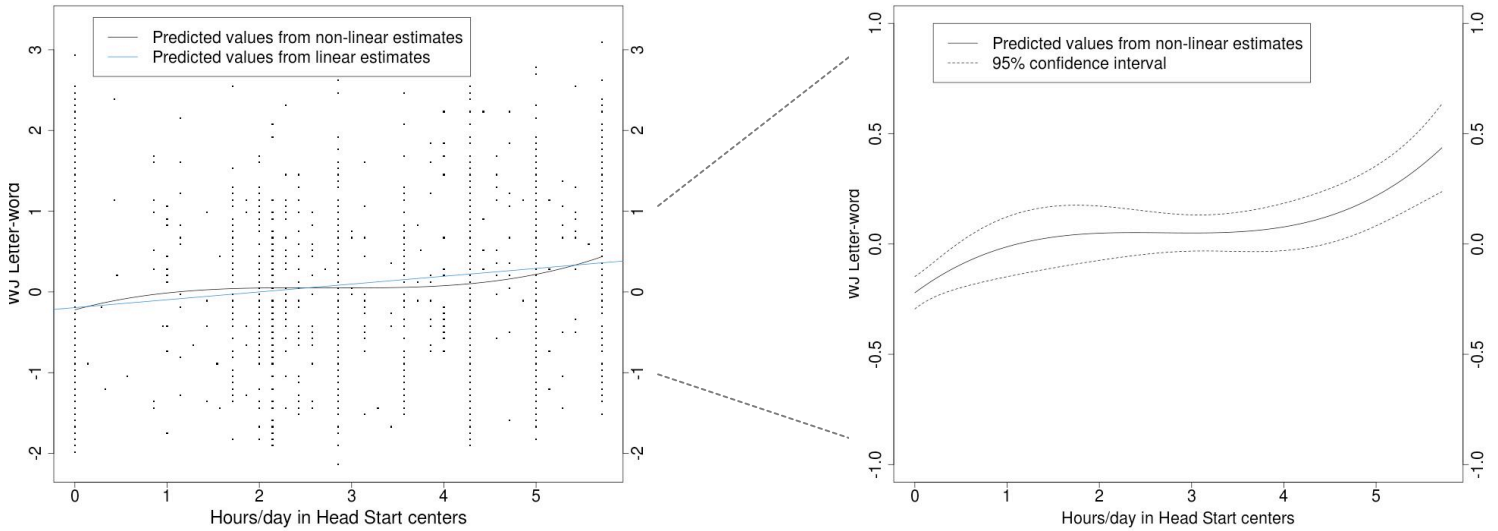

Figure 2b. Linear and non-linear effects of Head Start hours from Instrumental Variable approach: Age 4 cohort.