

**Abstract Title Page**  
*Not included in page count.*

**Title:** Is more time in Head Start always better for children? The moderating role of classroom quality

**Authors and Affiliations:**

Allison H. Friedman-Krauss  
New York University  
Department of Applied Psychology  
IES-Predoctoral Interdisciplinary Research Training Fellow

Maia C. Connors  
New York University  
Department of Applied Psychology  
IES-Predoctoral Interdisciplinary Research Training Fellow

Pamela A. Morris  
New York University  
Department of Applied Psychology

## **Abstract Body**

*Limit 4 pages single-spaced.*

### **Background / Context:**

*Description of prior research and its intellectual context.*

The 1998 reauthorization of Head Start called for a national evaluation of the Head Start program. The goal of Head Start is to improve the school readiness skills of low-income children. Yet characteristics of Head Start programs, such as their quality and the amount of time children spend in them may influence their effectiveness at achieving this goal. Previous research has demonstrated that children who spend more weekly hours in Head Start demonstrate larger cognitive gains than children who attend fewer weekly hours (Li, Farkas, Duncan, Vandell, & Burchinal, 2013). But the quality of care children attend during such hours may also matter, as there is mounting evidence regarding the importance of classroom quality for children's developing school readiness skills (Burchinal, Vandergrift, Pianta, & Mashburn, 2009; Zaslow et al., 2010). Thus, the effect of weekly hours in Head Start may vary by the quality of the Head Start classroom. Indeed, previous research has found that child care quality moderates the association between hours in child care and child behavior problems, but not cognitive skills (McCartney et al., 2010; Votruba-Drzal, Coley, & Chase-Lansdale, 2004).

### **Purpose / Objective / Research Question / Focus of Study:**

*Description of the focus of the research.*

The current study expands on previous research by using quasi-experimental methods that leverage the experimental context of the Head Start Impact Study (HSIS; Puma et al., 2010) to understand the extent to which Head Start classroom quality moderates the impact of weekly hours in Head Start on children's early language and math skills and externalizing behaviors. We begin by replicating Li and colleagues' (2013) instrumental variables (IV) analysis assessing the effects of weekly hours in Head Start on outcomes for children, leveraging the random assignment nature of the HSIS design and the "offer" of differing numbers of hours of Head Start in the treatment condition and zero hours of Head Start in the control condition. We then extend this work to account for the quality of the Head Start center children attend. We hypothesize that weekly hours in Head Start will be more strongly associated with outcomes for children enrolled in high quality programs as compared with children enrolled in low quality programs. In contrast to previous research that used samples of children enrolled in child care, the current study relies on a sample of children enrolled in educationally focused Head Start programs.

### **Setting:**

*Description of the research location.*

The Head Start Impact Study sample was designed to be nationally representative of 3- and 4-year-olds attending Head Start programs in the United States and included children in 22 states. Observations of classroom quality occurred in the child's Head Start center.

### **Population / Participants / Subjects:**

*Description of the participants in the study: who, how many, key features, or characteristics.*

This research uses data from the Head Start Impact Study in which 4,440 3- and 4-year-old

predominantly low-income children were randomly assigned off a waitlist to either receive an invitation to participate in Head Start services or to the control group. Children applied to 351 Head Start centers. Small centers were combined with nearby centers to form 202 center groups and children were randomly assigned to Head Start from these center groups rather than the Head Start center to which they applied. The sample for these analyses was limited to children who had data on at least one of the outcome measures, and to children in one randomly selected Head Start random assignment center per center group. We randomly selected one center in order that the assessment of quality be reflective of the center to which children applied (rather than the average quality across centers in center groups). After including only center groups in which there was non-missing data for children in both random assignment groups, the analytic sample included 2,482 children in 270 Head Start centers and 697 Head Start classrooms.

### **Intervention / Program / Practice:**

*Description of the intervention, program, or practice, including details of administration and duration.*

Children were randomly assigned to Head Start or to a control group. Children assigned to Head Start had access to the services that the Head Start program provided, including a certain number of hours of center-based care and related medical, dental, nutrition, and mental health services. The control group could not enroll in that Head Start center, but could enroll in other early care programs (i.e. non-Head Start center based programs and family child care) or were cared for at home by a parent. Children in the control group did not have access to Head Start but there were some “crossovers” in which children in the control condition attended other Head Start centers.

### **Research Design:**

*Description of the research design.*

Random assignment occurred prior to the beginning of the 2002-03 school year. Data collection began in the fall of 2002, after random assignment and continued through the spring of 2003.

### **Data Collection and Analysis:**

*Description of the methods for collecting and analyzing data.*

The current study utilizes data collected during the first year of the longitudinal Head Start Impact Study. Children’s early math and language skills were assessed using the Peabody Picture Vocabulary Test (PPVT; Dunn, Dunn, & Dunn, 1997) and Woodcock-Johnson Letter-Word Identification and Applied Problems subtests (Woodcock, McGrew, & Mather, 2001) during the fall and spring in the child’s home or Head Start center by trained data collectors. The PPVT measures children’s receptive vocabulary while the Letter-Word Identification task assesses children’s ability to identify and name letters and words. The Applied Problems tasks measures children’s ability to analyze and solve math problems. Children’s externalizing problems were assessed using a parent report based on the Child Behavior Checklist (Achenbach, Edelbrock, & Howel, 1987). Seven items regarding children’s hyperactive and aggressive behaviors were combined to form the externalizing problems scale (Alpha = 0.71). In order to improve the interpretation of findings, all outcome measures were standardized to z-scores.

Classroom quality was measured using the Early Childhood Environment Rating Scale (ECERS-R; Harms, Clifford, & Cryer, 1998) and the Arnett Caregiver Interaction Scale (CIS; Arnett,

1989) during the spring of 2003. The ECERS-R and CIS are observation tools used to measure early childhood classroom quality. Exploratory and confirmatory factor analysis were used to extract three construct specific factors across these two tools. The resulting construct specific factors are: Materials & Space for Learning, Positive Teacher-Child Interactions, and Negative Teacher-Child Interactions (Connors, Friedman-Krauss, Jones, & Morris, in preparation). In the spring of 2003 Head Start center directors reported the daily operating schedule for their center which was used to compute the number of weekly hours of Head Start offered. Parents reported the number of hours per week their child attended Head Start.

Our first step was to replicate previous work by Li and colleagues (2013) which, using an instrumental variables approach, found a positive effect of attending more weekly hours of Head Start on children’s math and language abilities and behavior problems. IV analyses allow for the approximation of a causal estimate by using an “instrument” to isolate the exogenous variance in a “mediator” variable (i.e. variance that is not correlated with child or family characteristics) and using that exogenous variation to predict the outcome of interest (Gennetian, Morris, Bos, & Bloom, 2005). IV analyses reduce the threat of selection bias due to characteristics of children and families that are associated with selection into Head Start care settings, quality of those care settings, and child outcomes. We used the number of hours per week *offered* by a Head Start center as an instrument; hours of Head Start *offered* was zero for any child randomized to the control group as they were not offered Head Start services. In the first stage we used hours of Head Start offered to predict the number of hours a child attended Head Start. In the second stage we estimated child outcomes from the predicted hours in Head Start generated from stage 1:

$$(1) Y_{Hours\ in\ Head\ Start} = \Pi_{10} + \Pi_{11}Hours\ Offered_{1i} + \Pi_{1k}Covariates_{ki} + \sum \Pi_{1n}CenterGroups_{ni} + \mu_{1i}$$

$$(2) Y_{child\ outcome} = B_0 + B_1Hours\ in\ HS + B_kCovariates_{ki} + \sum B_nCenterGroups_{ni} + \varepsilon_i$$

In both stages we included a set of indicators to control for the child’s center group and a limited set of baseline covariates (including children’s pre-random assignment outcomes) in order to increase the precision of the instrumental variable estimates (Gennetian et al., 2005). Multiple imputation was used to replace missing data on baseline covariates.

In a first phase of this work, we split the sample into groups of children attending high and low quality programs in order to test our hypothesis that weekly hours in Head Start is more strongly associated with positive child outcomes in high as compared with low quality centers. Children who were not in Head Start classrooms were missing a Head Start classroom quality score. In order to group these children into either high or low quality, we used the average Head Start classroom quality across all Head Start classrooms attended by children who applied initially to the same Head Start center. In doing so, we assume that the quality of children’s early care experiences is relatively homogenous within the neighborhood where they applied to Head Start. For each of the three construct specific quality factors we split the sample into high and low quality based on the median of the distribution of average quality. We repeated our analyses (equations 1 and 2) separately for each of the high and low quality groups and used *t* tests to compared the impact estimated in the high and low quality subsamples. Future analyses will consider quality and the interaction between the number of hours in Head Start and quality as endogenous variables in instrumental variables models, utilizing additional instruments.

## **Findings / Results:**

*Description of the main findings with specific details.*

Results of the IV analysis show that, first and foremost, the F-statistics from the first stage was greater than 30 and weekly hours of Head Start offered predicted weekly hours of Head Start attended ( $b=0.41$ ,  $S.E.=0.01$ ,  $p<0.001$ ) indicating sufficient strength in our instrument to isolate the exogenous variation in weekly hours of Head Start attended. Table 1 presents the impacts of weekly hours in Head Start on children's development. Children who attended more weekly hours of Head Start demonstrated significantly higher math and language skills at the end of the school year. Children who attended more weekly hours of Head Start were rated as having lower externalizing behaviors but this relationship was not significant (please insert table 1 here).

Tables 2, 3, and 4 present the results of the separate instrumental variables analyses for the high and low quality groups for each construct-specific quality factor. More hours in Head Start significantly predicted higher math in high but not low quality Head Start programs when quality was measured by Materials & Space or Negative Interactions. These differences in the effect of weekly hours in high and low quality Head Start were significant at the trend level. Additionally, the effect of weekly hours in Head Start on children's externalizing behaviors was statistically significantly different (at the trend level) in high quality compared to low quality Head Start when quality was measured by Materials & Space (please insert tables 2, 3, & 4 here).

## **Conclusions:**

*Description of conclusions, recommendations, and limitations based on findings.*

The results of this study are in line with previous findings that more hours in Head Start and higher quality classrooms are associated with higher school readiness skills. We find that both weekly hours in Head Start and the quality of the Head Start program are important, especially for children's math skills. Translating our hourly effects in high quality programs into full-day estimates (i.e., 40 hours) amounts to moderate effect sizes (ranging from 0.29 to 0.44) on math and language outcomes. These estimated effect sizes are in line with a recent meta-analysis of early childhood programs (Yoshikawa et al., 2013). Moreover, a child enrolled in a high quality, as measured by low negative interactions, full-day program is estimated to perform 0.32 standard deviations higher in math compared to a child enrolled in a low quality full-day program.

Notably, our analyses considered quality as exogenous (with the instrumental variables methodology only accounting for the endogeneity in weekly hours of Head Start). In future analyses, we plan to address this issue by instrumenting hours and quality of Head Start as well as their interaction in assessing their effects on outcomes for children.

These results suggest that program operating schedule and quality should be considered simultaneously when thinking about the circumstances under which Head Start has positive effects on children. As Head Start and other educationally focused early childhood programs expand, it may be that providing more hours is only be beneficial (and therefore cost effective) if programs provide and maintain high quality classrooms, and vice versa. The finding that more weekly hours in Head Start positively impacts children's academic skills highlights the tradeoff between the costs and benefits of enrolling more children for fewer hours or fewer children for more hours.

## Appendices

Not included in page count.

### Appendix A. References

References are to be in APA version 6 format.

- Arnett, J. (1989). *Arnett Caregiver Interaction Scale*. Retrieved from Jaeger, E. & Funk, S. (2001). *The Philadelphia Child Care Study: An Examination of Quality in Selected Early Education and Care Settings*. Philadelphia, PA: Saint Joseph's University.
- 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.
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A., (2009). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*. 25, 166-176.
- Connors, M. C., Friedman-Krauss, A. H., Jones, S. M., & Morris, P. A. (in preparation). Refining early measures of early childhood classroom quality.
- Dunn, L.M., Dunn, L.L., and Dunn, D.M. (1997). *Peabody picture and vocabulary test, third edition (PPVT)*. Circle Pines, MN: American Guidance Service.
- Gennetian, L., Morris, P.A., Bos, J., & Bloom, H. (2005). Using instrumental variables analysis to learn more from social policy experiments. In H. Bloom (Ed.) *Learning More from Social Experiments: Evolving Analytic Approaches*. New York: Russell Sage
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale (Revised Edition)*. New York, NY: Teachers College Press.
- Li, W., Farkas, G., Duncan, G.J., Vandell, D.L., & Burchinal, M. (2013). Effects of Head Start hours on children's cognitive, pre-academic, and behavioral outcomes: An instrumental variables analysis. Paper presented at the Society for Research on Educational Effectiveness.
- McCartney, K., Burchinal, M., Clarke-Stewart, K.A., Bub, K.L., Owen, M.T., Belsky, J., & the NICHD Early Child Care Research Network. (2011). Testing a series of causal propositions relating time spent in child care to children's externalizing behavior. *Developmental Psychology*, 46(1), 1-17. doi:10.1037/a0017886
- Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., ... & Spier, E. (2010). *Head Start Impact Study. Final Report*. Washington, D.C.: Administration for Children & Families.
- Woodcock, R.W., McGrew, K.S., and Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside Publishing.

- Votruba-Drzal, E., Coley, R.L., & Chase-Lansdale, P.L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, 75(1), 296-312. doi:10.1111/j.1467-8624.2004.00670.x
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Gormley, W., Ludwig, J., Magnuson, K.A., Phillips, D., & Zaslow, M.J. (2013). Investing in our future: The evidence base on preschool education. Ann Arbor, MI: Society for Research in Child Development and New York: Foundation for Child Development.
- Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L., & Burchinal, M. (2010). *Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature, OPRE 2011-5*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

## Appendix B. Tables and Figures

*Not included in page count.*

Table 1. IV estimates of the effects of hours per week of Head Start on child outcomes

	PPVT	WJ Letter Word	WJ Applied Problems	Externalizing Problems
<b>OLS</b>				
B <sub>1</sub> Hours	0.0024***	0.0071***	0.0037***	-0.0014
(se)	(0.0010)	(0.0012)	(0.0012)	(0.0013)
Covariates	xxx	xxx	xxx	xxx
F	18.82***	9.77***	8.05***	6.28***
<b>IV</b>				
B <sub>1</sub> Hours	0.0077***	0.0104***	0.0066***	-0.0005
(se)	(0.0016)	(0.0019)	(0.0020)	(0.0021)
Covariates	xxx	xxx	xxx	xxx
F	18.54***	9.64***	8.02***	6.27***

*Notes:* \*\*\* $p < 0.001$ . Dependent variables are standardized using z-scores. Control variables include baseline assessment of outcome, child gender, if the child was assessed at baseline in Spanish, if the child's mother had less than a high school diploma, an indicator for if the child applied to Head Start as a three- or four-year-old, and a set of indicator variables for the center group from which random assignment occurred to account for nesting.

Table 2. IV estimates of the effects of hours per week of Head Start on child outcome by high and low quality based on Materials & Space for Learning

	PPVT	WJ Letter Word	WJ Applied Problems	Externalizing Problems
<i>High Quality</i>				
B <sub>1</sub> Hours	0.0077***	0.0126***	0.0098***	-0.0036
(se)	(0.0021)	(0.0027)	(0.0028)	(0.0028)
covariates	xxx	xxx	xxx	xxx
F	19.55***	8.84***	7.33***	6.82***
<i>Low Quality</i>				
B <sub>1</sub> Hours	0.0072**	0.0079**	0.0023	0.0035
(se)	(0.0023)	(0.0028)	(0.0029)	(0.0031)
covariates	xxx	xxx	xxx	xxx
F	17.44***	10.23***	8.68***	5.63***
P value for Difference Between High and Low Quality	0.87	0.24	0.07 <sup>†</sup>	0.10 <sup>†</sup>

Notes: <sup>†</sup> $p < 0.10$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Dependent variables are standardized using z-scores. Control variables include baseline assessment of outcome, child gender, if the child was assessed at baseline in Spanish, if the child's mother had less than a high school diploma, an indicator for if the child applied to Head Start as a three- or four-year-old, and a set of indicator variables for the center group from which random assignment occurred to account for nesting.

Table 3. IV estimates of the effects of hours per week of Head Start on child outcome by high and low quality based on Positive Teacher-Child Interactions

	PPVT	WJ Letter Word	WJ Applied Problems	Externalizing Problems
<i>High Quality</i>				
B <sub>1</sub> Hours	0.0080***	0.0122***	0.0073*	-0.0011
(se)	(0.0023)	(0.0028)	(0.0029)	(0.0029)
covariates	xxx	xxx	xxx	xxx
F	16.27***	7.54***	6.35***	5.62***
<i>Low Quality</i>				
B <sub>1</sub> Hours	0.0070**	0.0083**	0.0047 <sup>†</sup>	0.0011
(se)	(0.0021)	(0.0027)	(0.0028)	(0.0029)
covariates	xxx	xxx	xxx	xxx
F	21.16***	11.71***	10.09***	6.94***
P value for Difference Between High and Low Quality	0.75	0.32	0.52	0.60

Notes: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Dependent variables are standardized using z-scores. Control variables include baseline assessment of outcome, child gender, if the child was assessed at baseline in Spanish, if the child's mother had less than a high school diploma, an indicator for if the child applied to Head Start as a three- or four-year-old, and a set of indicator variables for the center group from which random assignment occurred to account for nesting.

Table 4. IV estimates of the effects of hours per week of Head Start on child outcome by high and low quality based on Negative Teacher-Child Interactions

	PPVT	WJ Letter Word	WJ Applied Problems	Externalizing Problems
<i>High Quality</i>				
B <sub>1</sub> Hours	0.0085***	0.0079*	0.0109**	0.0012
(se)	(0.0025)	(0.0032)	(0.0034)	(0.0033)
covariates	xxx	xxx	xxx	xxx
F	18.43***	9.21***	6.42***	6.89***
<i>Low Quality</i>				
B <sub>1</sub> Hours	0.0069***	0.0119***	0.0029	-0.0009
(se)	(0.0020)	(0.0024)	(0.0024)	(0.0026)
covariates	xxx	xxx	xxx	xxx
F	18.40***	9.72***	9.69***	5.63***
P value for Difference Between High and Low Quality	0.62	0.33	0.06 <sup>†</sup>	0.62

Notes: <sup>†</sup> $p < 0.10$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Dependent variables are standardized using z-scores. Control variables include baseline assessment of outcome, child gender, if the child was assessed at baseline in Spanish, if the child's mother had less than a high school diploma, an indicator for if the child applied to Head Start as a three- or four-year-old, and a set of indicator variables for the center group from which random assignment occurred to account for nesting.