California Head Start Child Outcomes 2011: Methodology and Discussion of Results

Peter Huffaker, Partner
Child Care Results
April Morris, Partner
Child Care Results

February 2011

This document and the supporting bulletin were constructed through a creative collaboration of Child Care Results, the California Head Start Association, and the forty-two participating Head Start programs listed on the next page. The analysis was conducted by Child Care Results in the fall of 2010.

Thank-you to the California Head Start Association and the 42 California Head Starts and their teachers who partnered with us to carry out this study:

Anaheim City School District, Orange County
Berkley Albany YMCA, Alameda County
Casa Blanca Head Start, Riverside County
Center for Community and Family Services, Los Angeles County
Chicano Federation, San Diego County
Child Care Resource Center, Los Angeles County
Child Development Resources of Ventura County
Children of the Rainbow, San Diego County
City of La Habra Head Start, Orange County
Cochella Valley Unified School District, Riverside County
Community Action Commission of Santa Barbara, Santa Barbara County
Community Action Partnership of Kern, Kern County
Community Action Partnership of San Luis Obispo, San Luis Obispo County
Desert Sands Unified School District, Riverside County
Episcopal Community Services, San Diego County
Elk Grove Unified School District, Sacramento County
Hemet Unified School District, Riverside County
Irvine Unified School District, Orange County
Kidango, Santa Clara and Alameda Counties
KidZKount: Placer Community Action Council, Inc., Placer County
Lake Elsinore Unified School District, Riverside County
MAAC Project, San Diego County
Moreno Valley Unified School District, Riverside County
Neighborhood House Association, San Diego County
Orange County Head Start
Orange Children & Parents Together, Orange County
Palm Springs Unified School District, Riverside County
Pomona Unified School District, Los Angeles County
Riverside County Office of Education
Romoland School District, Riverside County
Sacramento City Unified School District, Sacramento County
Sacramento Employment & Training Agency, Sacramento County
San Diego Unified School District, San Diego County
San Jacinto Unified School District, Riverside County
San Juan Unified School District, Sacramento County
Santa Ana Unified School District, Orange County
Santa Cruz Community Counseling Center, Santa Cruz County
Sierra Cascade Family Opportunities, Lassen, Modoc, Plumas, & Sierra Counties
Tulare County Office of Education
Twin Rivers Unified School District, Sacramento County
University of Southern California Head Start, Los Angeles County
Women's Civic Improvement Club of Sacramento, Sacramento County

We owe additional thanks to the eight individuals who served on an advisory committee to shape the final product:

Peter Huffaker, Partner
Child Care Results
April Morris, Partner
Child Care Results
Rick Mockler, Executive Director
California Head Start Association
Kay Wernert, Director,
Marin Head Start
Wassy Tesfa, Head Start Administrator
Center for Community and Family Services
Judy Cashell, Child Development Administrator
Center for Community and Family Services
Jennifer McCulley, Education Manager
Sierra Cascade Family Opportunities
Brenda Poteete, Executive Director
Sierra Cascade Family Opportunities

This analysis of child assessment data on 24,500 children was conducted by Child Care Results in the fall of 2010. The 4 page bulletin can be found http://www.caheadstart.org/ChildOutcomes2011. An electronic version of this methodology can be found on-line at http://www.childcareresults.com/ChildOutcomes2011. The analysis was conducted by Child Care Results alone. Any errors are solely the responsibility of Child Care Results.

Child Care Results specializes in helping child care organizations get the most value from their child assessment, community assessment and survey data. Through careful analysis, we give decision makers the critical insights they need to make informed decisions. Our results are always presented in thoughtful and compelling visual formats to make it easy for everyone to understand and use the information. For more about Child Care Results, please visit www.childcareresults.com

The **California Head Start Association** is the unified voice providing leadership and advocacy for the Head Start community. The California Head Start Association is an important strategic partner representing Head Start interests in California and the nation. For more about the California Head Start Association, please visit http://www.caheadstart.org/

Introduction

Providing the Head Start experience for three-year-olds advances a child's development by an average of 4 to 8 months across all areas of development (effect sizes of 0.52 to 0.63). This finding is the result of a rigorous analysis of child assessment data from more than 24,500 children collected from 42 Head Start programs across California. Using Regression Discontinuity - a methodology widely used to evaluate preschool impacts - it is possible to isolate the impact of participating in the Head Start program as a three-year-old on child development. These findings provide direct evidence that California Head Start programs provide quality child development services and help children to be better prepared for Kindergarten.

Head Start is a federally funded comprehensive child development program for very low-income young children and their families. The primary target population for Head Start is 3 to 5 year-olds (and their families), while Early Head Start targets pregnant women and families with children 0 to 2 years-old. In 2009, California Head Start programs served more than 85,000 3 to 5 year-olds. Data from more than a quarter of those children was collected for this analysis.

Every year Head Start programs collect child assessment data on thousands of children. For only the second time, Child Care Results and the California Head Start Association worked together to consolidate this data and analyze it to determine the impact that California Head Start programs have on child development. The majority of California Head Start programs use the Desired Results Developmental Profile – Revised (DRDP-R), to track the development of children in their programs. Data from these assessments was used for this analysis.

About the DRDP-R

The Desired Results Developmental Profile– Revised (DRDP-R) was developed by the California Department of Education. The DRDP-R assessment for Preschoolers includes ten indicators. The indicators include fundamental areas of development (e.g. Math, Literacy, Social and Interpersonal Skills) and have been aligned to the Head Start outcome framework domains.

Table 1: Desired Results Indicators

Indicator	Description
Self Awareness & Self Concept	Children show self-awareness and a positive self-concept
Social Interpersonal Skills	Children demonstrate effective social and interpersonal skills
Self Regulation	Children demonstrate effective self-regulation in their behavior
Language	Children show growing abilities in communication and language
Learning	Children show interest, motivation, and persistence in their approaches to learning
Cognitive Competence	Children show cognitive competence and problem-solving skills through play and daily activities
Math	Children demonstrate competence in real-life mathematical concepts
Literacy	Children demonstrate emerging literacy skills
Motor Skills	Children demonstrate an increased proficiency in motor skills
Safety and Health	Children show an emerging awareness and practice of safe and healthy behavior

There are thirty-nine measures grouped into these ten indicators. Within each measure, children are observed and assessed at one of four developmental levels or as "Not yet at first level".

Table 2: Desired Results Preschool Developmental Levels

Developmental Levels *listed from least developed to most developed*
0 – Not Yet at First Level
1 – Exploring
2 – Developing
3 – Building
4 – Integrating

In validating the tool, researchers combined the indicators into six developmental domains (or indicator groupings). Since the validity of the tool was demonstrated using scores combined at the level of the domains, the analysis was conducted at that level and the Child Outcomes Bulletin reports on these six developmental domains.

Table 3: Desired Results Developmental Domains

Developmental Domains *sometimes known as Indicator Groupings*
1. Self Concept – Social Interpersonal Skills
2. Self Regulation – Safety and Health
3. Language – Literacy
4. Learning – Cognitive Competence
5. Math
6. Motor Skills

The analysis does not include the English language learners measures, added in the Head Start version of the DRDP-R. These measures specifically address the development of language skills for non-English speaking children but were not included in CDE studies validating the tool.

Sample Overview

Data from the DRDP-R was collected on 24,540 children enrolled in California Head Start programs in the fall of 2009. Data from 10,215 children were used in the primary analysis. Forty-two different Head Start programs contributed all of their preschool assessment data (for children 3 to 5 years-old). For analysis, the sample was divided into four different groups:

- Discarded Data – For simplicity, the analysis focuses on typically developing children (with no special needs) enrolled in center-based programs (not licensed family homes or home-based programs). Children for whom we were missing critical data elements (i.e. race, ethnicity, primary language, or gender), children outside the age range (i.e. less than three or more than five years-old), and children whose assessment dates were not between August and December 2009 were excluded from the data. The last reason that children were excluded is because they were three years-old and had been enrolled in Head Start the previous year. In total, data from

5,369 children were excluded from the analysis. Below is a breakdown of the number excluded by reason:

Table 4. Exclusion Reasons

Exclusion Reason	Number of Children
Not in a center-based program	1,695
Demographic data missing	3,245
Outside age range	57
Assessment not between Aug. and Dec. 2009	169
Children were 3 years-old and had prior enrollment	203

- Four Year-olds New to Head Start – In California, children are eligible for Kindergarten if they will be five years-old on December 2nd of the year of enrollment. 8,956 children from the sample would be eligible for Kindergarten in 2010 (for simplicity we call them four year-olds) and were newly enrolled to Head Start
- Four Year-olds Returning to Head Start – 4,584 children would be eligible for Kindergarten in 2010 (four year-olds) and were enrolled in Head Start in the prior year (as three year-olds) – they are returning for their second year of Head Start.
- Three Year-olds – 5,631 children would be eligible for Kindergarten in 2011 (for simplicity we call them three year-olds). All of the three year-olds were new to Head Start.

The analysis focuses on comparing 'Four Year-olds Returning to Head Start' and 'Three Year-olds'. 'Four Year-olds New to Head Start' are not included in the analysis, but do serve as an interesting comparison point. The table below identifies the characteristics of each group:

Table 5. Sample Demographic Breakdown

	Four Year-Olds Returning to Head Start	Three Year-Olds	Four Year-Olds New to Head Start
Total Number of Children	4,584	5,631	8,956
Female	53%	52%	50%
Male	47%	48%	50%
Latino-Spanish	50%	50%	52%
Latino-English	17%	14%	16%
African American-English	8%	9%	8%
White-English	8%	8%	7%
White-Other Language	4%	4%	3%
Multi Racial-English	3%	3%	2%
Multi Racial-Spanish	3%	3%	2%
Asian-Asian Language	3%	4%	4%
Asian-English	2%	2%	2%
Other	3%	3%	3%

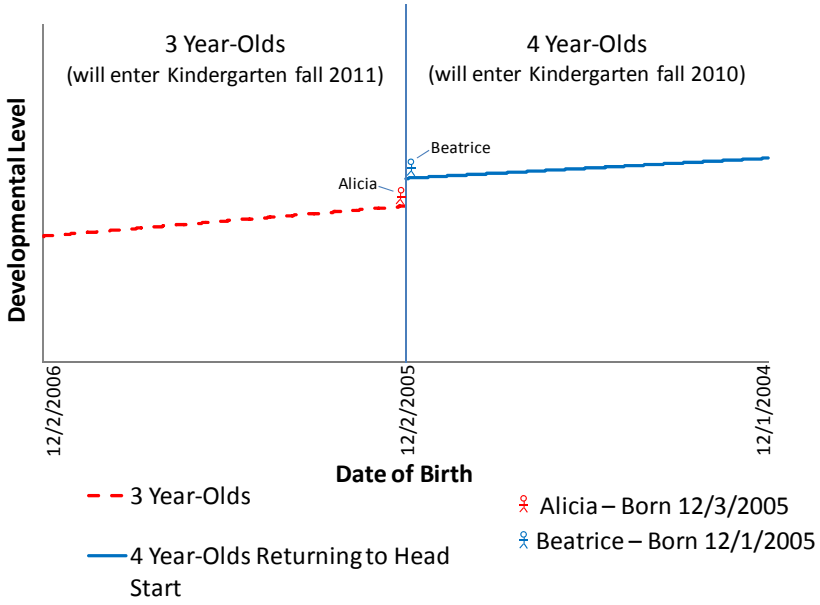
Study Design

To estimate the impact of participating in the Head Start program, the 'Four Year-Olds Returning to Head Start' were used as the experimental group and the 'Three Year-Olds' were used as the comparison group. The two groups were compared using a regression discontinuity (RD) design. Comparing these two groups controls for selection bias because both groups of children come from families who decided to enroll in Head Start as three year-olds and all of them meet the general Head Start eligibility criteria (typically family income below 100% of the poverty level, with rare exceptions for families with income between 100% and 130% of the poverty level). Which of the two groups a child falls into depends on which side of December 2, 2005 that the child was born.

The concept is easiest to understand by providing an extreme example. Consider two children: Alicia was born December 3, 2005 and Beatrice was born December 1, 2005. Born two days earlier, Beatrice will start Kindergarten in the fall of 2010, while Alicia will start Kindergarten in the fall of 2011. In the fall of 2009, Beatrice is starting her second year of Head Start while Alicia is starting her first. The biggest difference is that Beatrice was exposed to Head Start and Alicia was not. Beatrice (and her cohort) serves as the experimental group and Alicia (and her cohort) serves as the comparison group. We include children born further from the cutoff date by using a regression model to control for age.

The simplified diagram in Figure 1a illustrates the comparison using actual results from the Math Domain. The lines represent the typical or expected (best fit line) developmental level given a child's birth date. Both lines slope up to the right. This slope is simply an indication that older children are more developed than younger children. But at the cutoff point, there is a jump (discontinuity) in the line. This jump reflects the increased development that the children in the experimental group have due to participation in the program.

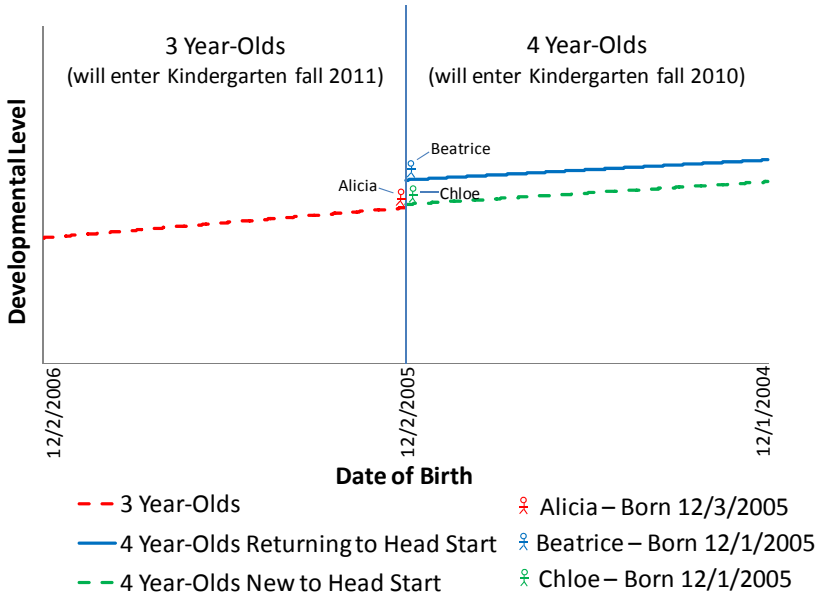
Figure 1a. Math
Expected Developmental Level by Date of Birth
From Fall 2009



As an interesting comparison point, we can include the results for the 'Four Year-Olds New to

Head Start' group (see Figure 1b). Chloe is our representative of the 'Four Year-olds New to Head Start' group. Like Beatrice, she was born on December 1st. But unlike Beatrice she did not attend Head Start last year. We can see that Beatrice's expected developmental level is much higher than Chloe's. This is likely due to Beatrice's participation in Head Start, but may be due to selection bias. The possibility of selection bias is why the analysis focuses on comparing Alicia to Beatrice ('3 Year-Olds' to '4 Year-Olds Returning to Head Start') rather than Beatrice to Chloe.

Figure 1b. Math
Expected Developmental Level by Date of Birth
From Fall 2009



Comparing Alicia to Chloe provides another interesting insight. The sharp eye will notice that Chloe is a head higher (developmentally) than Alicia. This suggests that children who enroll as three year-olds are generally less developed controlling for age than children who enroll as four year-olds. Generally Head Starts prioritize enrollment based on need. This analysis provides evidence that Head Start programs are successful in prioritizing needier (less developed) children as three year-olds.

To test the robustness of the findings, multiple statistical models were developed including all of the children with birth dates within 365 days on either side of the cutoff, within 180 days on either side of the cutoff, within 90 days on either side of the cutoff, and within 30 days on either side of the cutoff.

An additional complication in the data set is that the assessments were completed between August 2009 and December 2009. We were concerned that the diverse range of these dates could impact the results. We first attempted to control for the age of the child at the time of the assessment in the regression model. This proved difficult because the age at the time of the assessment was perfectly correlated to the age relative to the cutoff date.

Table 6. Month of Assessment

Month of Assessment	# of Children	% of Children
August 2009	99	1%
September 2009	255	2%
October 2009	4,132	40%
November 2009	5,662	55%
December 2009	67	1%

Since most children were assessed in November 2009, we ran an additional series of regressions looking at just children assessed in November 2009. The results presented in the California Head Start Child Outcomes Bulletin 2011 were based on the children within 90 days of the cutoff date and assessed in November 2009. These models were chosen because they were generally conservative (providing lower estimates for the impact of Head Start).

Regression Discontinuity has become the preferred methodology for determining the impact of preschool programs over the past ten years because it eliminates the problem of selection bias. Used with large samples, Regression Discontinuity is nearly as rigorous and reliable as a controlled experiment with random assignment.[1]

More commonly this methodology is used to evaluate preschool the year before Kindergarten (rather than two years prior) by using assessment data from children entering Kindergarten and those entering preschool. However, the Desired Results Developmental Profile-Revised (DRDP-R) is only used to assess preschoolers and not Kindergarteners. As a result, the analysis could only be used to evaluate Head Start's impact on three-year-olds. Please see Appendix 1 for a list of studies that use regression discontinuity design to evaluate the impact of preschool.

Potential Issues

Modeling: One disadvantage of regression discontinuity (RD) design is that accurate results depend upon accurate statistical modeling of the data. For this reason, RD incorporates careful graphical analysis and the construction of multiple statistical models to check for robustness. In total 48 (8 models for each of the six domains) separate models were developed using different aspects of the data. All of the models developed were linear. As noted in the graphic analysis section below, the relationships in the data were extremely linear. We believe that this linearity is related to the use of logits to convert the raw scores into scale scores. Regardless, quadratic and cubic models were not determined to be a good fit for this data.

There are two additional analyzes that could have been used to further validate the robustness of the findings. 1) Ideally, the analysis would have checked for robust standard errors, rather than merely standard errors. Software limitations prohibited the inclusion of robust standard errors.[2] 2) In addition

[1] More specifically, the internal validity of regression discontinuity are comparable to that of a controlled experiment with random assignment, although regression discontinuity requires larger sample sizes. See:
Berk, Richard. "Recent Perspectives on the Regression Discontinuity Design." Paper: University of Pennsylvania. (March 17, 2008). and
Black, Dan, Jose Galdo, and Jeffrey Smith. "Evaluating the Regression Discontinuity Design Using Experimental Data." Paper. (Version: April 2007).

[2] Statistica version 7.1 was used for the analysis. Yes, we promise to upgrade.

parametric models (such as linear regression), non-parametric models are often developed in RD. Due to time constraints non-parametric models were not used as part of this analysis. Although inclusion of these additional analyzes would be preferable, the clear patterns in the models used in this analysis leaves no doubt that they accurately portray the strength of the Head Start impact seen in this dataset.

Potential inter-rater reliability issues: The DRDP-R assessment results are based on individual teacher observations and assessments. Naturally, this raises the question of whether the assessments of two different teachers can be compared to one another. Research on the DRDP-R tool shows that it has high inter-rater reliability, between 0.87 and 0.90.[3] Still, it may be argued that inter-rater reliability would be lower in less controlled field settings where there may be variability in training, experience, or effort.

To the extent that inter-rater reliability issues do exist, we do not believe that would undermine the clear pattern of children with prior enrollment having higher levels of development. In order to disrupt this pattern, inter-rater reliability issues would have to disproportionately impact either three year-olds or 4 year-olds and disproportionately impact the developmental scores in one direction (either higher or lower). That such a specific pattern of inter-rater reliability problems would arise seems unlikely.

It is far more likely that any inter-rater reliability issues would be randomly distributed in the data. If that is the case the result would be an increase in the standard deviation which would lower the reported effect size. But a random distribution of errors would not impact the regression coefficients.

Graphical Analysis

Part of ensuring that the data are accurately modeled is a careful analysis of the data's distribution. The following 18 graphs show the scatter plots, linear plots, and lowess plots for each of the six developmental domains. The most important observation from these graphs is that all six domains follow a generally linear course.

A scatter plot simply marks the developmental level of each child against their birth date relative to the cutoff date (days from cutoff). The linear plot shows the best fit line relative to the scatter plot. The lowess plot or (locally weighted scatter plot smoothing) is similar to the best fit line, but it weights the effect of localized data points. This allows the best fit line to curve, allowing the observer to more easily identify non-linear relationships.

[3] http://www.wested.org/desiredresults/training/questions.htm 11/1/2009

Figure 2a. Self Concept - Social Interpersonal Skills
Scatterplot

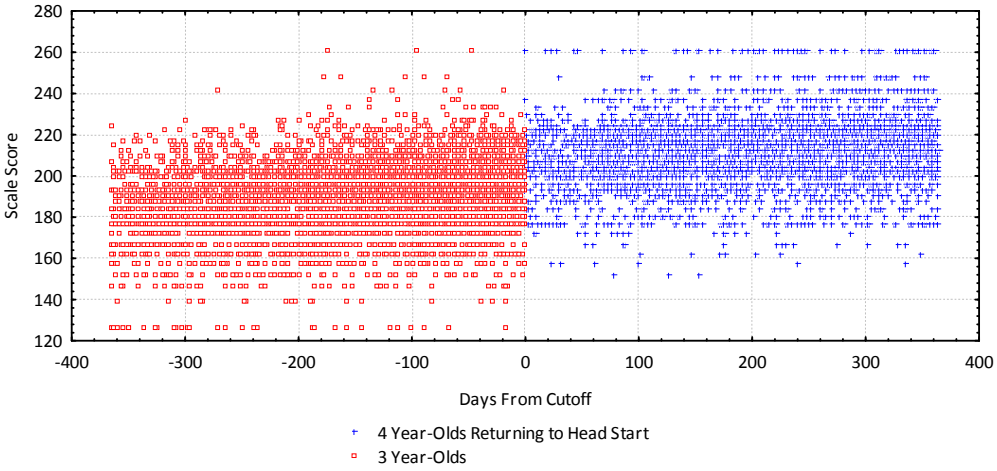


Figure 2b. Self Concept - Social Interpersonal Skills
Linear Plot

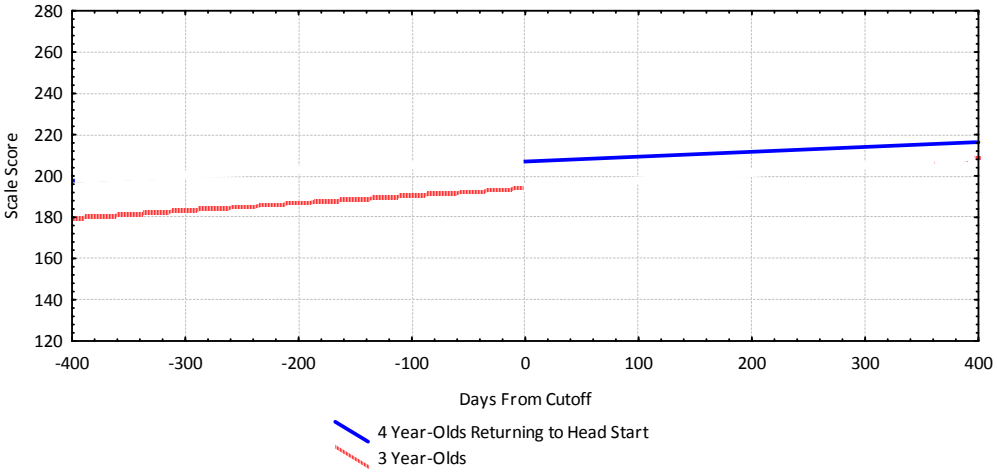


Figure 2c. Self Concept - Social Interpersonal Skills
Lowess Plot

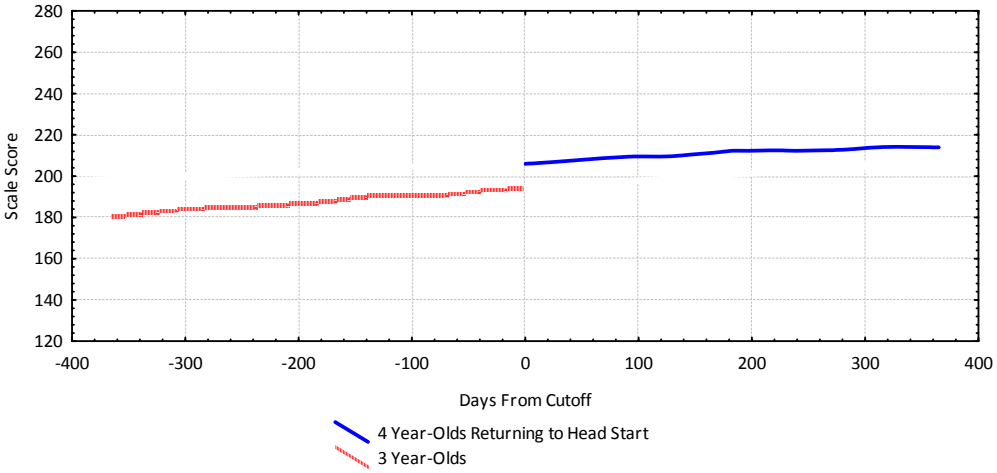


Figure 3a. Self Regulation - Safety and Health
Scatterplot

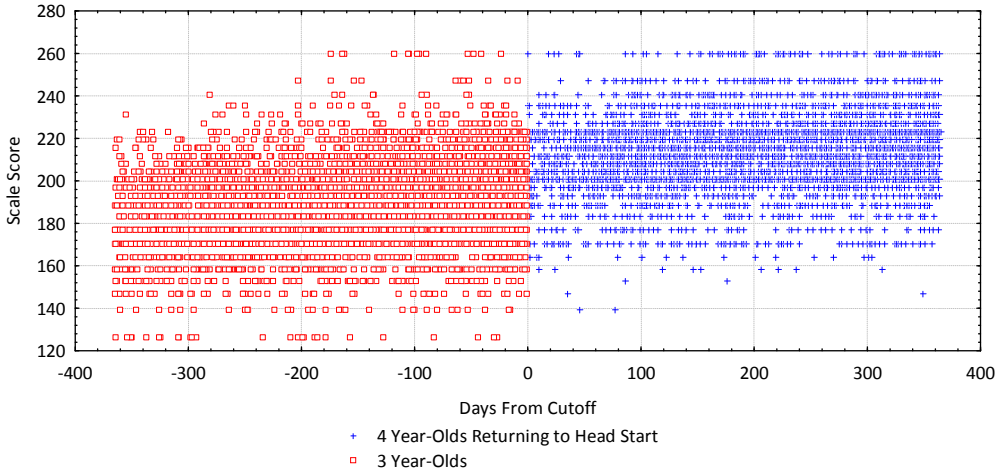


Figure 3b. Self Regulation - Safety and Health
Linear plot

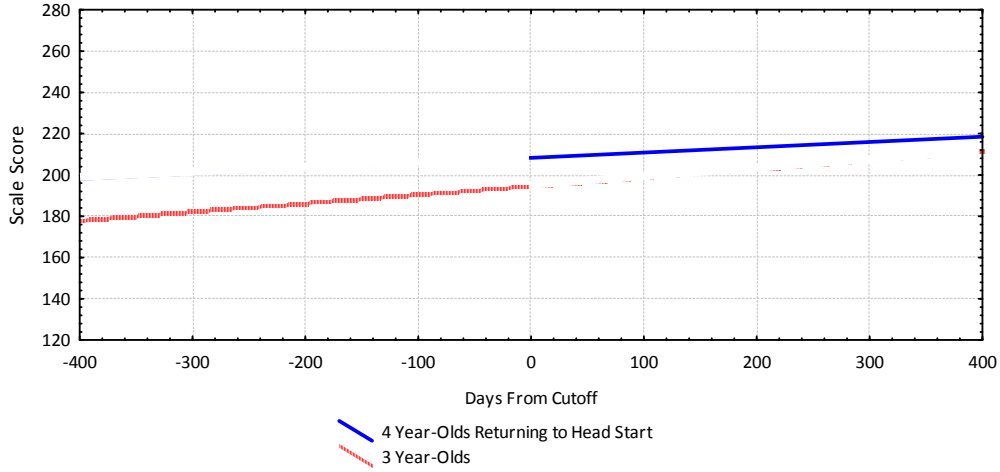


Figure 3c. Self Regulation - Safety and Health
Lowess Plot

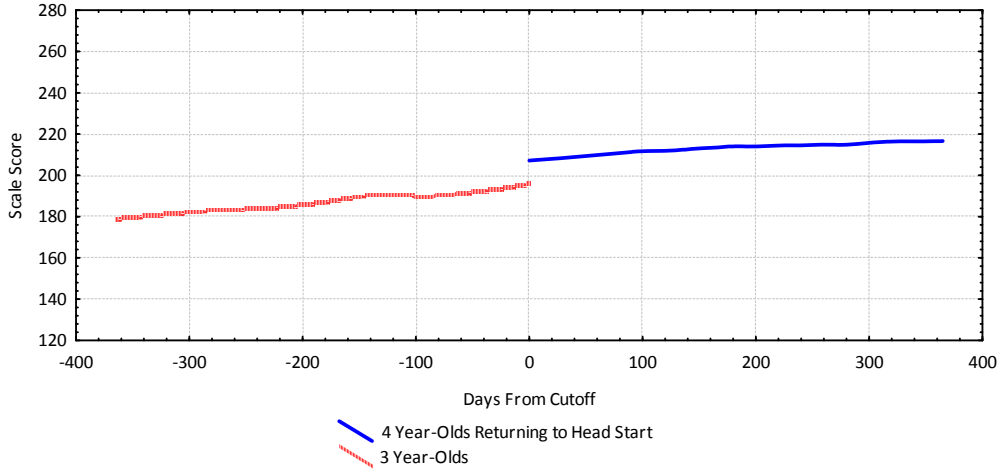


Figure 4a. Language - Literacy
Scatterplot

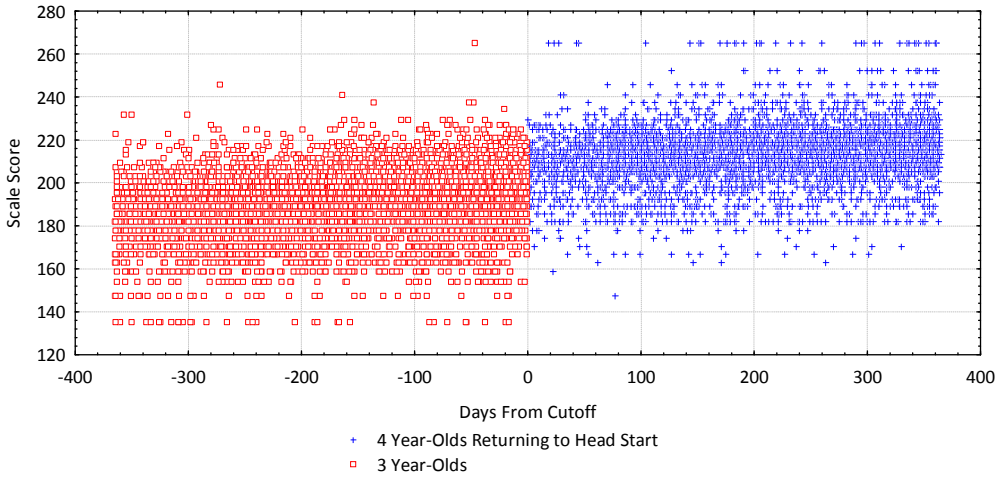


Figure 4b. Language - Literacy
Linear Plot

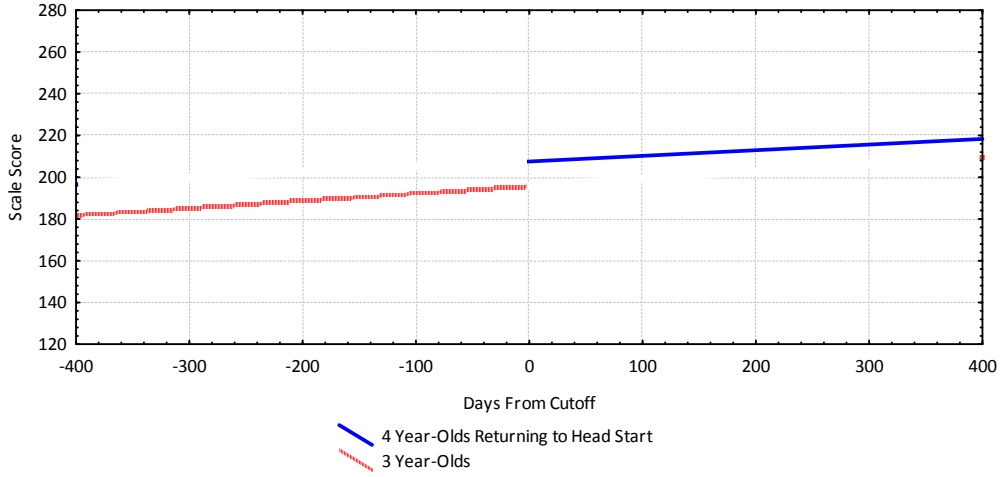


Figure 4c. Language - Literacy
Lowess Plot

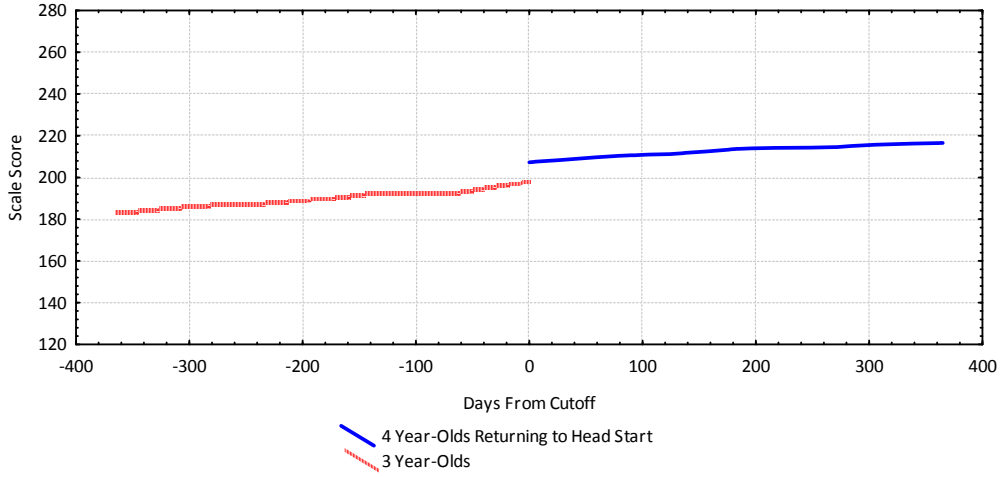


Figure 5a. Learning - Cognitive Competence
Scatterplot

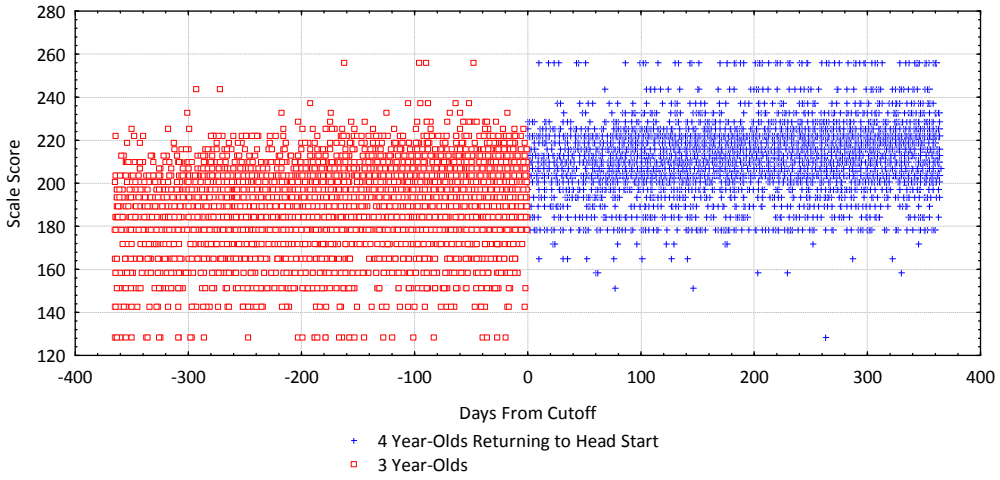


Figure 5b. Learning - Cognitive Competence
Linear Plot

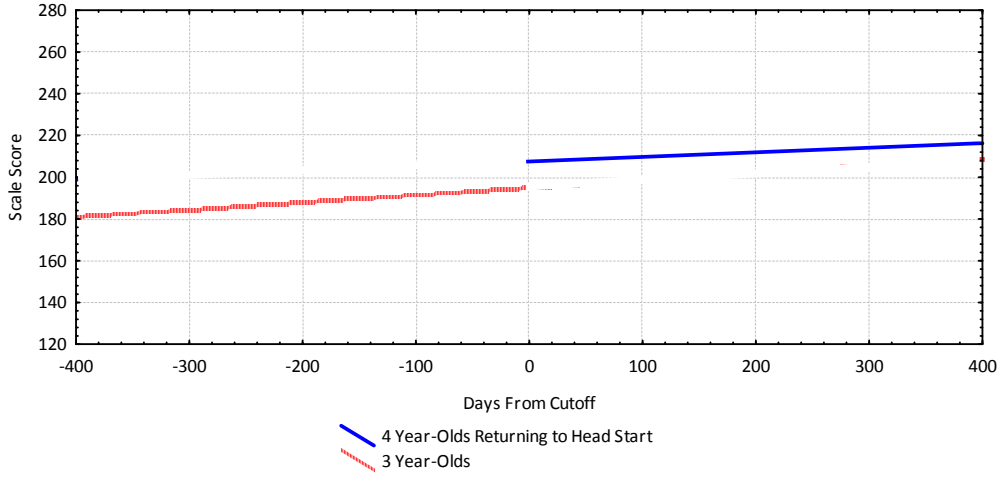


Figure 5c. Learning - Cognitive Competence
Lowess Plot

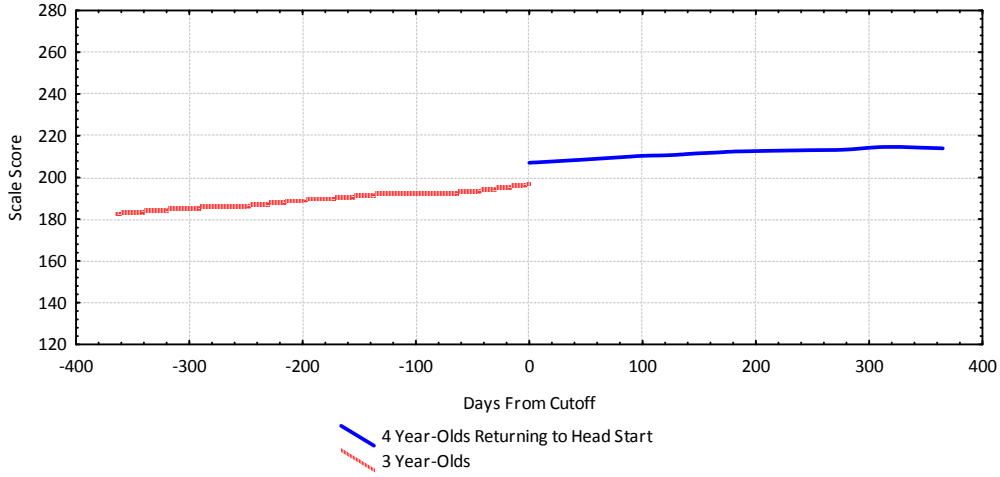


Figure 6a. Math
Scatterplot

Figure 6b. Math
Linear Plot

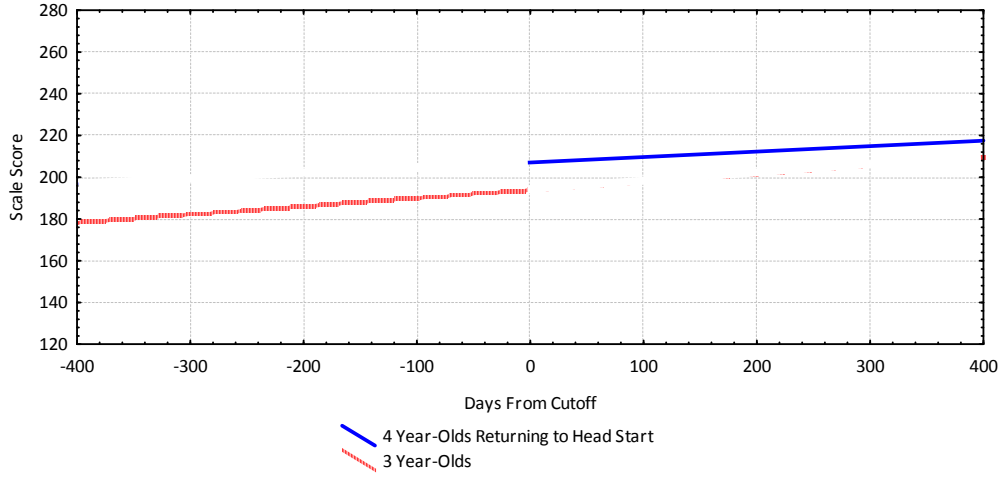


Figure 6c. Math
Lowess Plot

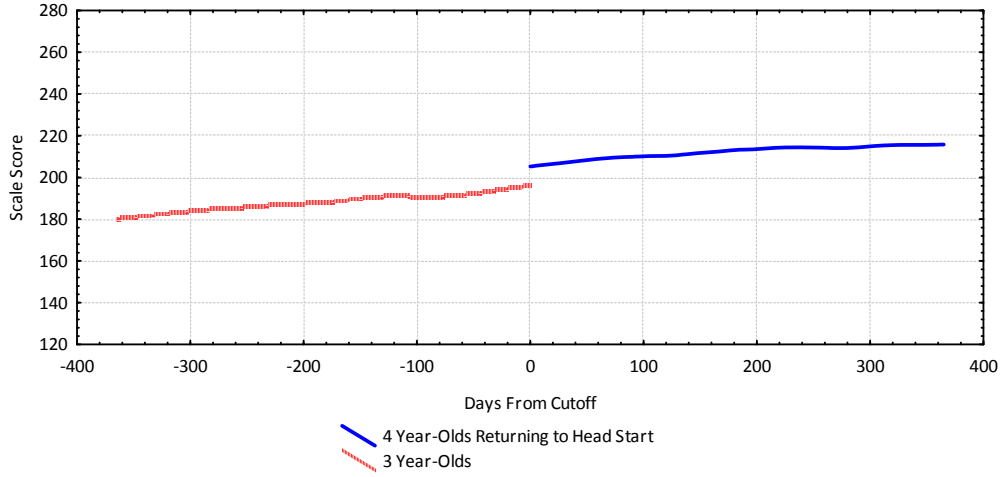


Figure 7a. Motor Skills
Scatterplot

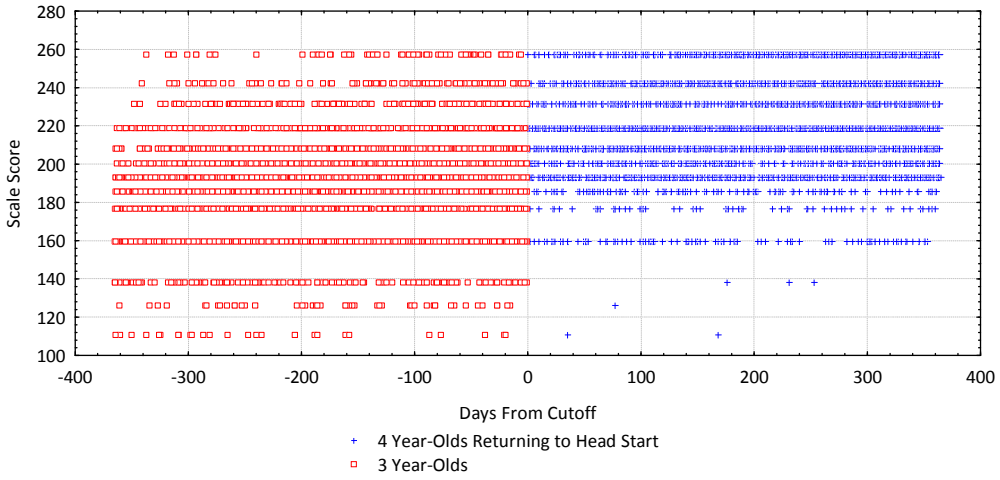


Figure 7b. Motor Skills
Linear Plot

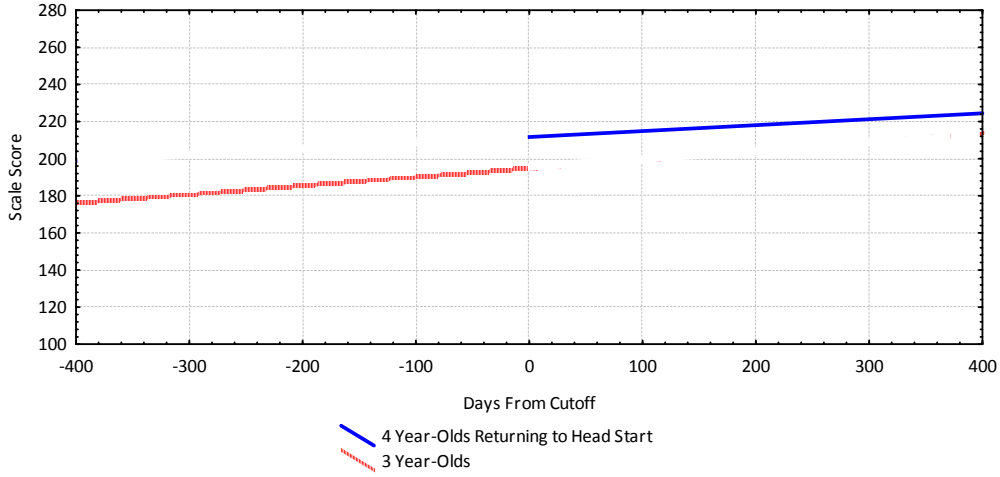
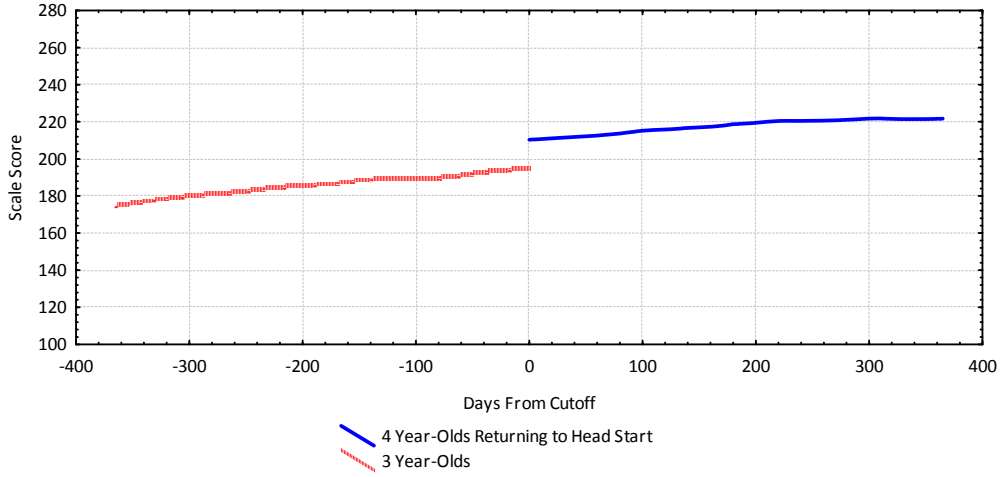


Figure 7c. Motor Skills
Lowess Plot



Regression Analysis

For each domain we developed eight regression models using different subsets of the data. Subsets were pulled out based on the month of the assessment and how far from the cutoff date the child was born. The dependent variable in each regression was the domain scale score. The independent variables included the number of days from the cutoff a child was born (age), gender, and race / ethnicity / primary language. A binary variable was included indicating whether a child was a '3 Year-Old' or a '4 Year-Old Returning to Head Start' – a one indicating that the child was a '4 Year-Old Returning to Head Start'. As a result, the coefficient for this variable is a measure of the jump or discontinuity that occurs at the cutoff. Most importantly it is a measure (in scale score points) of the impact of a child attending Head Start as a three year-old.

Table 7, reports the impact coefficients for all 48 models. Also reported are the significance of the coefficients, the standard errors and the sample size for each subset of the data. Except for a few of the smaller data sets, the results were significant at the 0.1% confidence level.

Generally the results across the different models were consistent. Although the coefficients were higher when data was included for 365 and 180 days around the cutoff. They were also higher when data from all the assessment months were included (rather than just assessments in November).

Table 7. Regression Scale Score Coefficients for Impact of Head Start

	All Children Assessed Aug. to Dec. 2009				Only Children Assessed in Nov. 2009			
Developmental Domain	365 Days	180 Days	90 Days	30 Days	365 Days	180 Days	90 Days	30 Days
Self Concept – Social Interpersonal Skills	12.71*	12.46*	11.60*	11.67*	11.30*	11.89*	**10.72***	11.86*
	(0.69)	(0.97)	(1.37)	(2.40)	(0.94)	(1.34)	**(1.94)**	(3.50)
	n = 10,032	n = 5,145	n = 2,546	n = 847	n = 5,563	n = 2,827	**n = 1,403**	n = 449
Self Regulation – Safety and Health	13.71*	13.35*	11.94*	11.45*	12.23*	12.49*	**10.81***	14.36*
	(0.77)	(1.10)	(1.58)	(2.80)	(1.07)	(1.55)	**(2.27)**	(4.14)
	n = 10,036	n = 5,160	n = 2,552	n = 847	n = 5,579	n = 2,838	**n = 1,408**	n = 449
Language – Literacy	11.48*	10.98*	9.10*	8.38*	10.28*	10.15*	**8.08***	8.56**
	(0.60)	(0.85)	(1.23)	(3.91)	(0.83)	(1.18)	**(1.73)**	(3.09)
	n = 10,002	n = 5,132	n = 2,545	n = 846	n = 5,553	n = 2,824	**n = 1,406**	n = 449
Learning – Cognitive Competence	12.51*	11.96*	10.99*	8.94*	11.27*	11.42*	**10.32***	10.26**
	(0.66)	(0.93)	(1.33)	(2.30)	(0.91)	(1.29)	**(1.87)**	(3.36)
	n = 10,064	n = 5,162	n = 2,557	n = 849	n = 5,576	n = 2,835	**n = 1,412**	n = 450
Math	12.79*	12.17*	10.77*	8.54*	11.31*	10.95*	**10.21***	10.62**
	(0.64)	(0.90)	(1.28)	(2.23)	(0.88)	(1.24)	**(1.80)**	(3.37)
	n = 9,998	n = 5,124	n = 2,536	n = 840	n = 5,538	n = 2,815	**n = 1,397**	n = 447
Motor Skills	16.80*	16.15*	15.04*	15.75*	16.44*	16.17*	**13.22***	13.96**
	(1.00)	(1.42)	(2.04)	(3.52)	(1.38)	(1.96)	**(2.85)**	(5.12)
	n = 10,136	n = 5,194	n = 2,570	n = 855	n = 5,625	n = 2,857	**n = 1,420**	n = 454

* significant at 0.1% confidence level.

** significant at 1% confidence level.

Standard errors in parentheses. Note these are not robust standard errors.

The results published in the California Head Start Child Outcomes Bulletin 2011 are in bold.

Effect Size

The effect size was calculated by dividing the impact coefficient by the standard deviation of the comparison group ('Three Year-Olds'). The lower range of the effect size calculations (used for the California Head Start Child Outcomes Bulletin 2011) were between 0.52 and 0.63. This compares quite favorably to the findings of the 2005 National Head Start Impact Study. That study found the treatment on the treated effect size (which is what this report analyzed) for three year-olds to be between 0.15 to 0.35.[4] Table 9 includes the effect sizes from a range of other studies. The first three studies in that table were evaluating the effect of preschool on 4 year-olds. In all of these other studies, the children were assessed using a tool other than the DRDP-R and one must be careful when comparing effect size across different tools.

Table 8. Effect Size Calculations

	All Children Assessed Aug. to Dec. 2009				Only Children Assessed in Nov. 2009			
Developmental Domain	365 Days	180 Days	90 Days	30 Days	365 Days	180 Days	90 Days	30 Days
Self Concept – Social Interpersonal Skills	0.74	0.73	0.68	0.70	0.64	0.68	**0.62**	0.70
Self Regulation – Safety and Health	0.70	0.68	0.60	0.59	0.60	0.61	**0.53**	0.71
Language – Literacy	0.75	0.72	0.59	0.56	0.65	0.66	**0.52**	0.57
Learning – Cognitive Competence	0.73	0.71	0.65	0.56	0.64	0.66	**0.60**	0.63
Math	0.77	0.76	0.67	0.56	0.66	0.67	**0.63**	0.65
Motor Skills	0.67	0.65	0.59	0.65	0.64	0.64	**0.52**	0.56

The results published in the California Head Start Child Outcomes Bulletin 2011 are in bold.

Table 9. Effect Size Calculations from other Studies

Project	Effect Size Range	Length of Preschool
Head Start National Study	0.147 to 0.319	one year of preschool
Tulsa Head Start Program	0.334 to 0.514	one year of preschool
Tulsa Public School Pre-K Program	0.355 to 0.985	one year of preschool
Abecedarian Project	1.08 average	three years of preschool
Perry Preschool Study	0.77 to 1.16	two years of preschool

Gormley, Jr., William T.; Phillips, Deborah; Gayer, Ted "Preschool Programs Can Boost School Readiness" Georgetown University. http://www.crocus.georgetown.edu/reports/scilong.pdf 12/01/09.

Communicating the Results – Age Comparisons

Equally important to performing careful and accurate analysis is communicating the results in a clear and valid form. Our target audience for this analysis is both the Head Start Community and state and federal policy makers. For an academic audience, scale score coefficients and effect size would be sufficient (those are the language of academia). But for policy makers and practitioners these measures are too abstract and too unfamiliar. The search for a more concrete and tangible expression of the

[4] Ludwig, Jens and Phillips, Deborah. "The Benefits and Costs of Head Start" *Social Policy Report.* Volume XXI, Number 3. (2007).

results settled on the differences in ages at which children reach a particular benchmark. These ages were calculated by solving the regression equation (for each model) for the age at which a child reaches the developing level with and without Head Start experience. The demographic variables were set to the overall sample averages so that the resulting ages reflected a 'typical' child in the sample. The impact coefficient were alternated between 0 and 1 to simulate participation in Head Start.

The age calculations were heavily influenced by the coefficients of the 'days from cutoff' variable. For the models that only used data within 30 days of the cutoff, these coefficients were a bit erratic due to the limited range of birth dates in the data set. For this reason the age calculations for these models were deemed unreliable and not reported.

Although not as precise as the scale score coefficients or effect sizes, the age calculations communicate the same message without exaggerating or under-reporting the impact – while making the impact more tangible. Table 10 reports the age calculations for 36 of the regression models.

Table 10. Age Calculations
Age at which Regression Models Predict Children will reach the
Developing developmental level with and without Head Start Experience

		All Children Assessed Aug. to Dec. 2009				Only Children Assessed in Nov. 2009			
Developmental Domain	Age Category	365 Days	180 Days	90 Days	30 Days	365 Days	180 Days	90 Days	30 Days
Self Concept – Social Interpersonal Skills	Without Head Start	4.56	4.75	4.52	Age Calculations based on 30 day regressions not reliable	4.45	4.74	**4.45**	Age Calculations based on 30 day regressions not reliable
	With Head Start	3.40	3.61	3.75		3.56	3.64	**3.80**	
	Age Difference	1.15	1.14	0.77		0.89	1.10	**0.66**	
Self Regulation – Safety and Health	Without Head Start	4.39	4.53	4.37		4.30	4.49	**4.31**	
	With Head Start	3.17	3.45	3.67		3.32	3.51	**3.72**	
	Age Difference	1.23	1.08	0.70		0.98	0.98	**0.59**	
Language – Literacy	Without Head Start	4.85	4.94	4.53		4.73	4.90	**4.47**	
	With Head Start	3.99	4.05	4.09		4.05	4.09	**4.11**	
	Age Difference	0.86	0.89	0.44		0.68	0.80	**0.36**	
Learning – Cognitive Competence	Without Head Start	4.64	4.80	4.55		4.53	4.80	**4.53**	
	With Head Start	3.50	3.74	3.85		3.64	3.79	**3.89**	
	Age Difference	1.14	1.06	0.70		0.89	1.02	**0.64**	
Math	Without Head Start	4.90	5.05	4.69		4.74	4.92	**4.72**	
	With Head Start	4.03	4.10	4.11		4.08	4.14	**4.13**	
	Age Difference	0.87	0.95	0.57		0.66	0.78	**0.58**	
Motor Skills	Without Head Start	3.91	3.92	3.93		3.89	3.93	**3.92**	
	With Head Start	2.44	2.88	3.14		2.40	2.87	**3.44**	
	Age Difference	1.47	1.05	0.79		1.49	1.05	**0.48**	

All ages are in years.

The results published in the California Head Start Child Outcomes Bulletin 2011 are in bold.

Concerns over data use and accountability

In the wake of the national No Child Left Behind legislation, many in the early childhood education (ECE) community fear that assessment data (or DRDP-R data specifically) will evolve to be used as a tool for accountability. Or more bluntly from the perspective of many in the ECE field – there is a fear that DRDP-R data will someday be used to punish programs and teachers. Some may feel that this analysis is a step in that direction.

We recognize that there is some risk that this type of analysis (using DRDP-R data to evaluate program effect) may be misused within an accountability framework. However, there is also tremendous value in giving programs and policy makers a direct measure of program effect. Ultimately, we do not believe that it makes sense to ignore the value that can be gleaned from DRDP-R data because of fears that the

data may be misused.

We believe strongly that the DRDP-R tool is not appropriate for use as an accountability tool. As an observation based tool, the DRDP-R is subjective by nature. Using incentives or pressure on programs or teachers to meet certain benchmarks could undermine the integrity of the results and jeopardize the usefulness of assessment data. The DRDP-R tool and DRDP-R data can be very useful for guiding program activities and understanding program results. But to protect that value, DRDP-R outcomes should not be used for accountability. (This does not apply to making programs accountable for adequate implementation of the DRDP-R process).

Conclusion

The largest contribution that this analysis makes is to demonstrate that it is possible to use operational or administrative data to effectively evaluate program outcomes for Head Start. Such an approach comes with many constraints. For example, the choice of how children are assessed is limited to the assessment tools chosen by the program operators. In California, the California Department of Education's development and use of the Desired Results Development Profile-Revised, which has been adopted by most Head Start programs in the state, provides a large swath of consistency across organizations. Still many Head Start's use other assessment tools, so it is still not possible to develop a fully universal or representative sample of California Head Start programs using the DRDP-R. The DRDP-R is also quite different from the assessment tools most commonly used to evaluate preschool programs. This inconsistency with other evaluations makes it difficult to compare the findings to that of other outcome studies. Perhaps most importantly, using operational data means that researchers cannot control and ensure the quality of the data collected.

Significant advantages to using operational data provide a counter weight to its disadvantages. Using operational data is much more cost effective than typical outcome studies. There is less disruption to programs, sample sizes are much larger, and the evaluations can be part of on-going reporting rather than periodic one-time events. At a time when both federal and state policy makers are looking for ineffective programs reduce or eliminate, this analysis provides strong evidence of the effectiveness of the Head Start programs included in this evaluation.

Ultimately, this type of analysis of operational data should serve as a supplement to the more traditional and expensive evaluation methods.[5] Such a combination would give policy makers a much clearer picture than the dearth of information currently available on program impact allows.

[5] A great example of a more traditional research approach can be found in this unpublished assessment of the California State Preschool Program. Information was not available on the costs of the research. Barnett, Steven W., Carollee Howes, and Kwanghee Jung. "California's State Preschool Program: Quality and Effects on Children's Cognitive Abilities at Kindergarten Entry." Final Report to the First 5 California Children and Families Commission. (March 25, 2009).

Appendix 1. Bibliography

List of other studies using regression discontinuity design to assess preschool:

Barnett, Steven W., Carollee Howes, and Kwanghee Jung. "California's State Preschool Program: Quality and Effects on Children's Cognitive Abilities at Kindergarten Entry." Final Report to the First 5 California Children and Families Commission. (March 25, 2009).

Gormley Jr., William, Deborah Phillips, and Ted Gayer. "Preschool Programs Can Boost School Readiness." Paper: Georgetown University.

Hustedt, Jason T., W. Steven Barnett, Kwanghee Jung, and Alexandra Figueras. "Impacts of New Mexico PreK on Children's School Readiness at Kindergarten Entry: Results from the Second Year of a Growing Initiative." Paper, The National Institute for Early Education Research: Rutgers University. (June 2008)

Lamy, Cynthia, W. Steven Barnett, and Kwanghee Jung. "The Effects of Oklahoma's Early Childhood Four-Year-Old Program on Young Children's School Readiness." Paper, The National Institute for Early Education Research: Rutgers University. (December 2005).

Ludwig, Jens and Douglas L. Miller. "Does Head Start Improve Children's Life Chances? Evidence From a Regression Discontinuity Design." Paper, Georgetown University, University of California, Davis (2006).

Wong, Vivian C., Thomas D. Cook, W. Steven Barnett, and Kwanghee Jung. "An Effectiveness-Based Evaluation of Five State Pre-Kindergarten Programs." *Journal of Policy Analysis and Management.* Vol. 27, No.1. 122-154. (2008).