Experimental evidence on the distributional effects of Head Start

Hilary Hoynes (UC Berkeley)

Joint with
Marianne Bitler (UC Irvine) and
Thad Domina (UC Irvine)

April 2015

Overview

- There is substantial evidence that human capital investment in the first years of life generates high returns.
- Public preschools are prominent in current policy discussions and there is an active state movement in this area
- Currently, the most important *public* program in this area is Head Start.
- In this paper, we take advantage of nationally representative RCT on Head Start and comprehensively evaluate the program's impacts on cognitive and non-cognitive outcomes.
- We focus on estimating the distributional effects of the program

Motivation and Context

1. Head Start Impact Study
   - First national randomized experiment of the program
   - The existing research relies largely on non-experimental identification strategies
     - ❖ Family / sibling fixed effects estimators
   - Given the changing environment in public early childhood education, as well as the changing demographics of children in the U.S., evidence on current effects of the program is valuable

Motivation and Context (cont.)

2. Distributional Approach
   - We examine how the impacts of the program vary across subgroups: race, ethnicity, English language learners
   - We use quantile treatment effects to explore how the impacts of the program vary across the distribution of skills.
     - Do the gains accrue disproportionately to the higher or lower end of the skill distribution?
   - These approaches allow us to test the alternative theories about effects of early education: compensatory role vs. skills-beget-skills

Motivation and Context (cont.)

3. "Explaining" large short run gains in Perry preschool
   - Perry preschool experiment (and other model preschools in the 1960s) generated large short run gains to IQ; to some extent these have not been replicated in the HS setting: here we explore what might explain those effects:
     - Perry was targeted at very disadvantaged population
     - Counterfactual care setting in the 1960s was different (many center options today; very little in the 1960s)
     - [Very intensive intervention]

Motivation and Context (cont.)

- An additional motivation is the finding in Head Start that the positive effects on test scores fade-out in early elementary school (Currie and Thomas 1995, Garces et al 2002) yet there are positive long term effects (Ludwig and Miller 2007, Deming 2009)
- Evidence for 'fade-out and rebound' has also been found for class size reduction (Dynarski et al 2012)
- This provides a context for our analysis of HSIS and our focus on the heterogeneous effects.

Our approach given this context

- We comprehensively examine the effects of HS on cognitive outcomes during the preschool years. Test for the presence of heterogeneous effects of the program:
  - Effects across the distribution using quantile treatment effects
  - Effects across demographic groups (race/ethnicity, language), baseline cognitive skills
  - ➢ By looking across groups, we explore how much the Perry results derive from its disadvantaged population
- We use variation across subgroups in control group's care setting to explore the role of the counterfactual care setting

Our approach given this context (cont)

- Examine how the gains to HS change as children enter elementary school.
  - Is there evidence of fadeout? Are there persistent effects in some subgroups or in some parts of the distribution?
- What about impacts on non-cognitive outcomes? Do they exist and are they persistent?
- To address incomplete take-up and crossovers, we use IV for our mean impacts and QTE

Background on Head Start and early child education in the U.S. context

Background on Head Start

- Goal: improve social and cognitive outcomes for low-income pre-school children
- Promotes school readiness for low-income children, primarily ages 3 and 4.
- Started as part of War on Poverty in 1965
- Primarily funded federally, but locally administered.
- Currently serving around 900,000 children at $7 billion

Background on Head Start (cont.)

- Eligibility – family must have income below poverty line (or receive public assistance TANF/SSI).
- (10 percent of slots must go to disabled)
- Services: education and cognitive development, health care, nutrition & social services
- Run by 1,600 local agencies (non-profits, government, church, school systems)
- Local programs have considerable discretion in the design and operation of programs → much variation in programs, quality
- Obama Administration implementing new HS evaluation programs: requiring re-competition for centers not performing above standard

HS is a large share of Federal early childhood $

Program	Federal $2010, billions
Head Start (no Early HS)	6.70
Early Head Start	1.00
Child care subsidies	5.70
Child Care Food Program	1.40
Child Care Tax Credits	2.20
DoD child care	0.75
Title 1 preschool	0.50
Preschool special education	0.57
Infant/toddler disability interventions	0.63
Home visits	0.10

Source: Haskins and Barnett (2010).

Demonstrates that HS is the most important piece of federal $$ in this area.

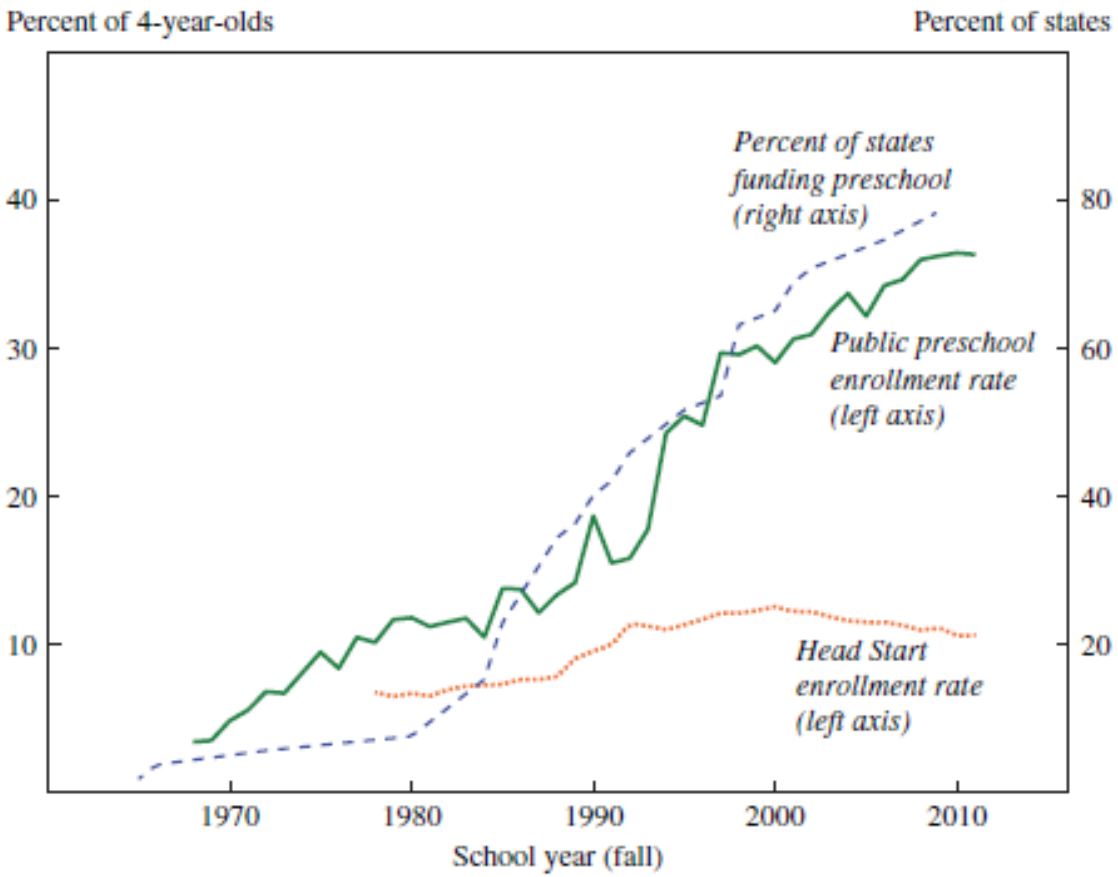
Head Start is an important preschool program for low-income children

	All Kids	Income in Bottom 20%	Income in Top 20%
3-year olds			
Head Start	8	20	1
Special Education	4	1	10
Private	36	22	68
Other Public	3	3	3
Any Preschool	51	45	82
4-year olds			
Head Start	13	29	0
Special Education	6	2	4
Private	42	22	75
Other Public	13	11	12
Any Preschool	74	64	90

Source: Haskins and Barnett (2010), from 2005 NHES.

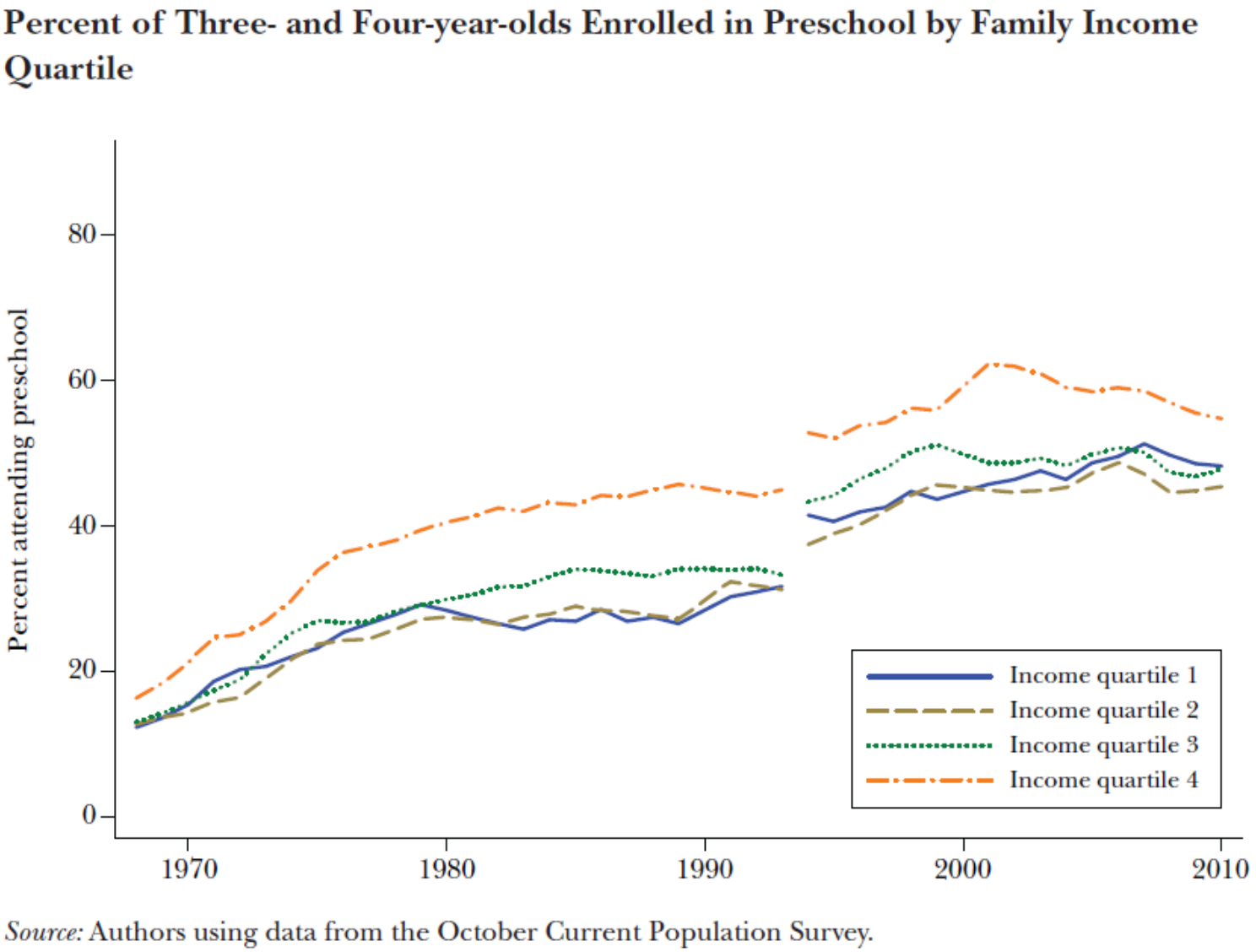
BUT, state public programs are growing in importance (especially for 4 year olds)

Figure 1. Percent of 4-Year-Olds Enrolled in Public Preschool Programs and Percent of States Funding Preschool Programs, 1965–2011



Sources: October CPS (public preschool enrollment rate), Head Start Bureau (numerator of Head Start enrollment rate), Vital Statistics (denominator of Head Start enrollment rate), and National Institute for Early Education Research (state funding dates).

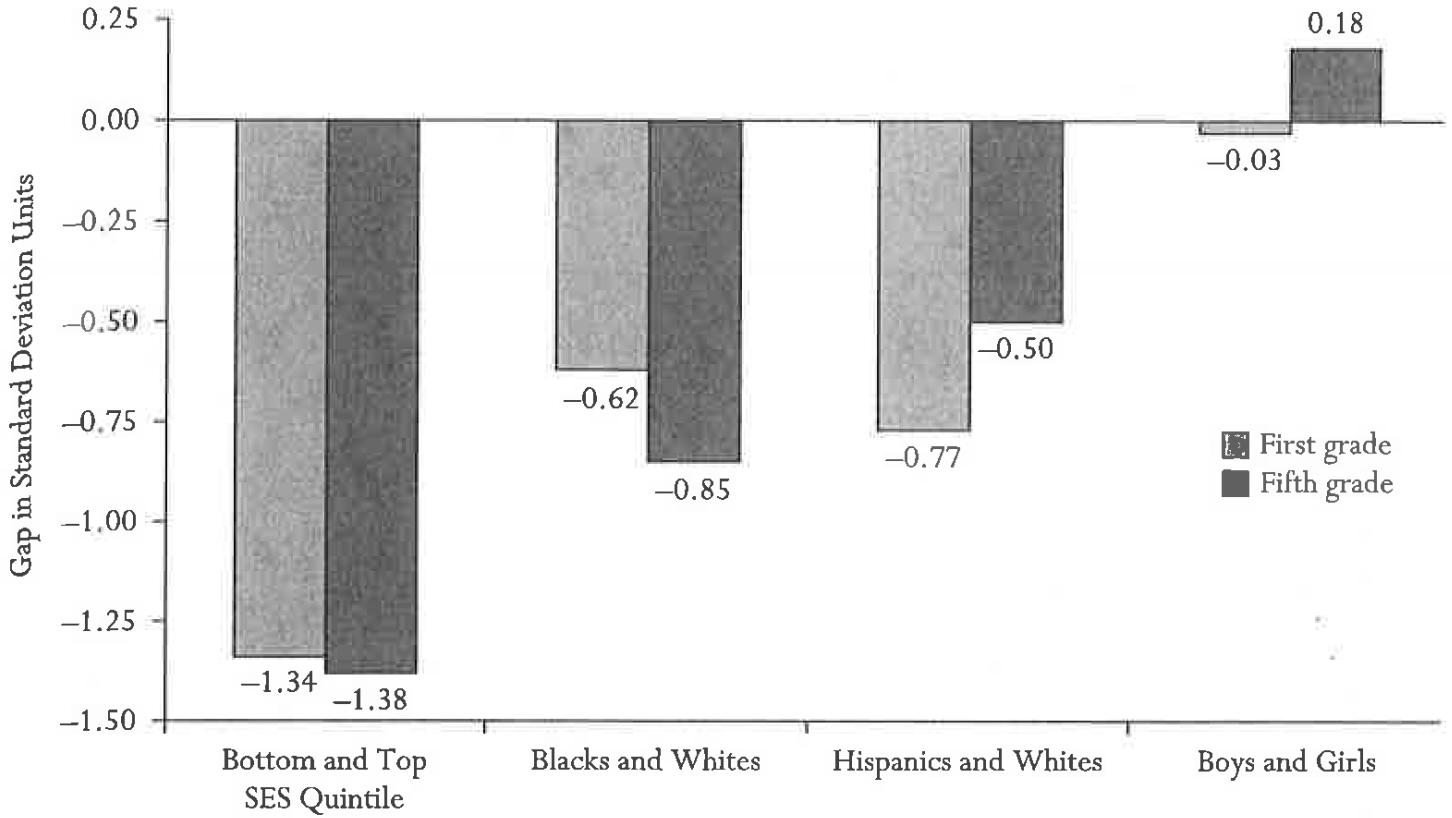
Why you should care: Disparities in preschool enrollment



Source: Greg Duncan and Katherine Magnuson "Investing in Preschool Programs," *Journal of Economic Perspectives*, 2013.

Why you should care: SES disparities are large

FIGURE 3.2 *Math Gaps in Kindergarten and Fifth Grade*



Source: Authors' calculations based on Early Childhood Longitudinal Study, Kindergarten Cohort (National Center for Education Statistics n.d.).

Source: Greg Duncan and Katherine Magnuson "The Nature and Impact of Early Achievement Skills, Attention and Behavior Problems," in *Whither Opportunity*, Duncan and Murnane, 2011.

Prior literature and theoretical expectations

Existing evidence on Head Start

- Effects on cognitive skills: Several studies use sibling / family fixed effects design and find positive effects on test scores but they fade out (Currie and Thomas 1995, Garces, Currie & Thomas 2002)
- Longer term outcomes: Consistent evidence using different quasi experimental designs that despite the fadeout in test scores, there are positive effects on longer term health, education and labor market outcomes
  - Deming (2009) & Garces et al (2002) sibling / family fixed effects; Miller & Ludwig (2007) RD on program rollout; Carneiro & Ginja (2014) RD on income eligibility rules
- Using the HSIS, Gelber and Isen (2011) find improvements in parental involvement (some human capital enhancing) that persist. Kline and Walters (2014) and Feller et al (2014) explore heterogeneity in impacts due to the counterfactual care setting.

Other related evidence

- Evidence from intensive pre-school interventions shows significant long term effects:
  - Perry preschool, Abecedarian (Heckman et al 2010, Anderson 2008)

Expectations: What should we expect in terms of heterogeneous effects of Head Start?

- *Compensatory role:* Largest effects of an education intervention accrue to those with the lowest skills. This has been found for early childhood education in observational studies (e.g. Magnuson 2004, Duncan and Magnuson 2013)
  - Head Start performance standards explicitly focus attention on basic skills and the program's mission is to serve at risk children → this may be relevant here
- *Skill begets skill / dynamic complementarities:* Skills students bring into early child education correlate with gains → gains to higher achieving students (Cunha & Heckman 2010)
  - Aizer and Cunha (2012) show largest effects and less fade out of Head Start for those with most early human capital

- These alternative theories provide motivation for our approach, in examining impacts across the distribution and across groups

Head Start Impact Study

- Mandated by Congress in 1998 (as part of reauthorization); goal to assess effects on school readiness and parental outcomes.
- First randomized experimental evaluation of HS
- Randomly assigned 3- and 4-year old children at oversubscribed programs to either an offer of placement in a HS program or no offer → "lottery design"
- Treatment is offer for slot for one year
- All students are first time applicants
- Sample: 2,449 in 3-year cohort, 1,993 in 4-year cohort
- 84 HS regional programs; randomization at 353 centers

Head Start Impact Study (cont)

- Timeline:
  - Spring 2002: Applications, lottery
  - Fall 2002: Head Start (treatment) year begins, [baseline testing and survey]
  - Spring 2003: Head Start (treatment) year ends
  - Followed through 3rd grade (spring 2008 for 3-year cohort and spring 2007 for 4-year cohort)
  - 3-year cohort potentially has 2nd HS year (not explicit component of HSIS)

Data available for outcomes through 1st grade

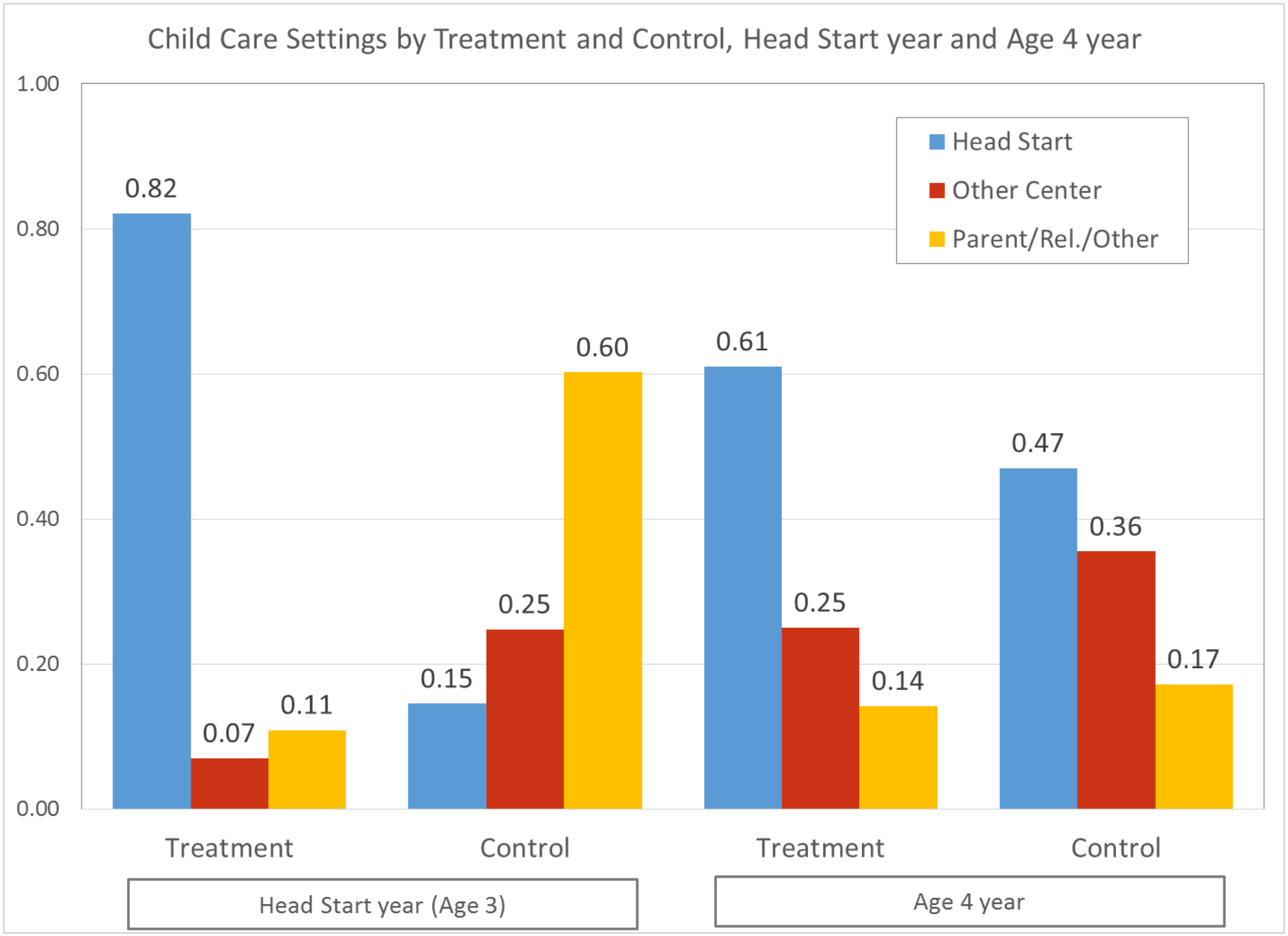
Our analysis sample: 3-year old cohort

- Eligibility for HSIS required children to be first time HS applicants
  - Not binding for 3-year olds, more so for 4-year olds; result is that 4 year olds look more disadvantaged
- Given current policy setting (e.g., the growth of public pre-K programs) 4-year olds have more options other than HS. There are fewer options for 3-year olds.
  - Cascio and Schanzenbach (2013) review and analyze public pre-K programs. 40 states had programs by 2011 (although only a few are universal).

Our sample consists of 2,449 3-year olds, 1464 in treatment and 985 in control

HSIS treatment and counterfactual

- As is common for this type of intervention, the offer of treatment does not translate one-for-one into Head Start participation
- There were "no-shows": about 16 percent of treatment children did not end up at a HS center
- There were "crossovers": about 15 percent of control children ended up in HS (presumably a different center than application)
- Given this, we show intent-to-treat (the impact of being offered a slot) but focus on treatment-on-treated (effect of HS using offer as instrument)
- Additionally, many consider the counterfactual care setting to be an important determinant to the estimated treatment effects of a early childhood intervention; we explore this


Child Care Settings by Treatment and Control, Head Start year and Age 4 year
1.00
0.80
0.60
0.40
0.20
0.00
Head Start
Other Center
Parent/Rel./Other
0.82
0.07
0.11
0.15
0.25
0.60
0.61
0.25
0.14
0.47
0.36
0.17
Treatment
Control
Treatment
Control
Head Start year (Age 3)
Age 4 year

Balance test:

Demographics well balanced at baseline

Demographics also indicate that HS applicants are a relatively disadvantaged group

	Baseline Child Weights	
	Control Mean	Difference T-C
Child characteristics		
Non-Hispanic white	0.344	-0.022
Non-Hispanic black	0.338	0.005
Hispanic	0.318	0.017
Female	0.527	-0.023
Spanish speaker at home	0.257	0.011
Low risk	0.789	-0.041*
Medium risk	0.156	0.013
High risk	0.055	0.028*
Special needs	0.103	0.031**
Lives with both biological parents	0.499	-0.003
Urban	0.793	0.001
Mother/caregiver characteristics		
Non-Hispanic white	0.364	-0.020
Non-Hispanic black	0.339	0.001
Hispanic	0.298	0.018
Mother was teenager at birth	0.176	-0.043*
High School dropout	0.346	-0.032
Only high school diploma/GED	0.325	0.033
More than high school	0.329	-0.001
Mom is married	0.457	-0.018
Mom is divorced	0.149	-0.010
Mom is never married	0.394	0.028
Age 20–24	0.316	-0.052**
Age 25–29	0.332	-0.016
Age 30–39	0.259	0.029
Age 40 or older	0.064	0.038***

Balance Test for HSIS (cont)

- There is imbalance in the timing of the tests (later for C) and in whether they took the test (more missing tests for C). [Missing baseline data is imputed by HSIS]
- Also, not all children were assessed prior to school year; so baseline test could capture some of treatment

	Control mean	Unadjusted T-C
Fall 2002 test month		
Before November	0.16	0.099***
November	0.32	0.029
After November	0.26	-0.013
No fall assessment (imputed)	0.26	-0.116***

Unadjusted, uses baseline weights.

Inverse propensity score weights

- We control for baseline characteristics in our estimates using inverse propensity score weights.
  - Deals with imbalance in month of baseline assessment and (relatively minor imbalance in Xs).
- Useful approach given our use of QTE (Firpo 2007)
- Estimate probability of T using logit with baseline sampling weights – predict probability of treatment $\hat{p}_i$
- Form inverse propensity score weight:

$$\hat{w}_i = \frac{T_i}{\hat{p}_i} + \frac{1 - T_i}{1 - \hat{p}_i}$$

- Control for baseline demographics, deciles for baseline PPVT for each of the four assessment month groupings, dummies for center of RA

	Baseline Child Weights		Inv. P-Score Weights
	Control Mean	Difference T-C	Difference T-C
Child characteristics			
Non-Hispanic white	0.344	-0.022	-0.001
Non-Hispanic black	0.338	0.005	0.001
Hispanic	0.318	0.017	0.0002
Female	0.527	-0.023	0.017
Spanish speaker at home	0.257	0.011	0.016
Low risk	0.789	-0.041*	0.044**
Medium risk	0.156	0.013	-0.019
High risk	0.055	0.028*	-0.025*
Special needs	0.103	0.031**	0.008
Lives with both biological parents	0.499	-0.003	0.024
Urban	0.793	0.001	0.006
Mother/caregiver characteristics			
Non-Hispanic white	0.364	-0.020	-0.008
Non-Hispanic black	0.339	0.001	0.010
Hispanic	0.298	0.018	-0.001
Mother was teenager at birth	0.176	-0.043*	0.023
High School dropout	0.346	-0.032	0.001
Only high school diploma/GED	0.325	0.033	-0.024
More than high school	0.329	-0.001	0.023
Mom is married	0.457	-0.018	0.022
Mom is divorced	0.149	-0.010	0.017
Mom is never married	0.394	0.028	-0.038
Age 20–24	0.316	-0.052**	0.036
Age 25–29	0.332	-0.016	0.021
Age 30–39	0.259	0.029	-0.016
Age 40 or older	0.064	0.038***	-0.042***
Fall 2002 test month/no Fall assessment			
Before November	0.162	0.099***	0.015
November	0.318	0.029	-0.022
After November	0.260	-0.013	-0.025
No Fall asssessment (imputed)	0.259	-0.116***	0.031

The First Stage

- It is not completely obvious what should be the first stage
- Our position is that participation in Head Start is the appropriate first stage.
  - We use this for the first stage for all outcomes in all years of the HSIS
- An alternative might be “any center” or “number of years in HS”
- Of course, any smaller first stage than we use implies that our IV estimates (reported here) are too low.

First Stage (Table 2)

	Treatment Mean	Control Mean	Difference T-C
Head Start in HS Year			
Head Start (Administrative report)	0.857	0.153	0.705***
Parent Report, Spring 2003			
Head Start	0.823	0.146	0.677***
Other Center	0.068	0.252	-0.183***
Family day care	0.014	0.064	-0.050***
Parent/relative	0.094	0.536	-0.442***
Not reported	0.001	0.002	-0.001
Parent Report, Spring 2004			
Head Start	0.608	0.473	0.135***
Other Center	0.250	0.355	-0.105***
Family day care	0.018	0.015	0.003
Parent/relative	0.077	0.103	-0.025
Kindergarten	0.016	0.021	-0.005
Not reported	0.031	0.033	-0.002

Outcomes measured in HSIS

Cognitive Skills

1. PPVT: measures vocabulary and verbal ability
2. Woodcock Johnson III Pre-academic Skills (Applied problems, Letter/Word, Spelling): more "achievement" oriented

- We focus on these tests because they are available for each year. Also, PPVT and early math assessments (WJIII Applied Problems) are known to be predictive of later life achievement (Duncan et al 2007)

Outcomes measured in HSIS

Non-cognitive skills

1. Adjustment Scales for Preschool Intervention (ASPI): emotional, behavioral adjustments
    - 144 items; collapsed to 5 behavioral dimensions and 3 situational dimensions. We also collapse further to 3 standardized indices
2. PIANTA scale: measures closeness, conflict and positive-ness of relationship

Limitation: Teacher reports not available until Kindergarten (come from parent reports).

Mean impacts of HSIS on PPVT

	Inverse P-Score Weights			Baseline Child Weights	
	Control mean [SD]	Reduced form (SE)	2SLS (SE)	Control mean [SD]	Reduced form (SE)
PPVT scores					
Baseline PPVT, fall 2002	231	-0.003		231	-0.88
	[38]	(1.84)		[39]	(2.25)
PPVT, spring 2003	251	7.20***	10.20***	252	6.56***
	[38]	(1.64)	(2.40)	[37]	(2.04)
PPVT, spring 2004	298	2.89	4.15	299	2.49
	[40]	(1.81)	(2.60)	[41]	(2.31)
PPVT, spring 2005	339	0.21	0.30	340	0.76
	[29]	(1.29)	(1.84)	[29]	(1.49)
PPVT, spring 2006	358	2.00	2.90	358	3.05
	[30]	(1.42)	(2.07)	[30]	(1.94)

- No significant difference at baseline (good)
- Gains in preschool year [ITT: 0.2 SD, IV: 0.3 SD or 40% incr.]
- Mean impacts not much affected by inverse propensity score weighting
- Fade-out (positive, not significant) in elementary school

	Inverse P-Score Weights		
	Control mean [SD]	Reduced form (SE)	2SLS (SE)
PPVT scores			
Baseline PPVT, fall 2002	231	-0.003	
	[38]	(1.84)	
PPVT, spring 2003	251	7.20***	10.20***
	[38]	(1.64)	(2.40)
PPVT, spring 2004	298	2.89	4.15
	[40]	(1.81)	(2.60)
PPVT, spring 2005	339	0.21	0.30
	[29]	(1.29)	(1.84)
PPVT, spring 2006	358	2.00	2.90
	[30]	(1.42)	(2.07)
PPVT missing or imputed			
No baseline PPVT or imputed	0.222	0.031	
	[0.416]	(0.022)	
No spring 2003 PPVT	0.190	-0.009	
	[0.393]	(0.023)	
No spring 2004 PPVT	0.179	0.009	
	[0.383]	(0.022)	
No spring 2005 PPVT	0.233	0.005	
	[0.423]	(0.022)	
No spring 2006 PPVT	0.251	0.001	
	[0.434]	(0.022)	

Additionally, with inverse p-score weights there is good balance in missing scores

RESULT 1:

Effect of HSIS on cognitive scores during the Head Start year for the full 3-year old sample

Quantile Treatment Effects

- Why look at the distribution?
  - Substantive question: where are the gains of early childhood interventions? High or low in the skill distribution?
  - Nonparametric way to explore the theoretical predictions
  - By focusing on means we could miss important effects (Bitler, Gelbach and Hoynes 2006)
  - Allow for testing for compensatory versus skill-beget-skills
- QTE examines impacts on *ex post* cognitive skills. An alternative is to estimate impact across distribution of baseline skills. We choose QTE because:
  - Not all children are assessed at baseline, C children are assessed later than T children & for some the baseline is several months after start of treatment → baseline test not ideal

Quantile treatment effect

$$\Delta_q = y_q(1) - y_q(0)$$

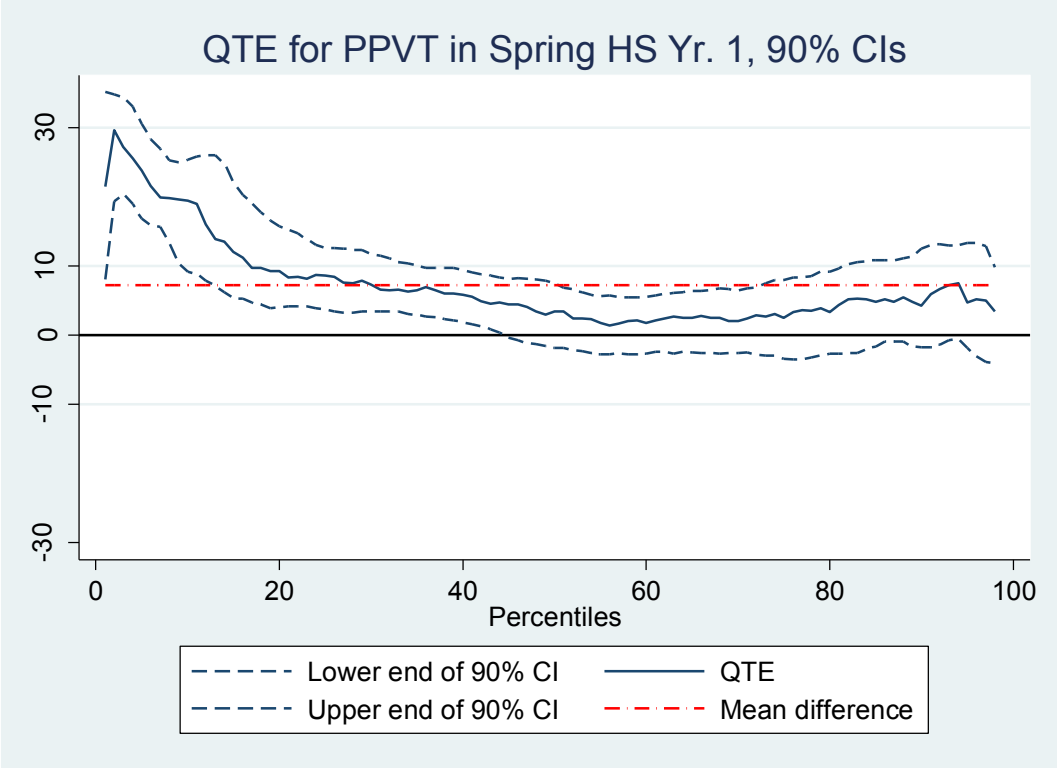
- $y_q(t)$ = *qth* quantile of the marginal distributions (of T and C)
- QTE is the difference between quantiles of the T and C distribution
- For example, the QTE at percentile 50 is the difference in medians of the T and C distribution.
- *Interpretation: change in expected value of the outcome at the qth quantile when the offer is made to the control distribution.*
- Given random assignment: the impact of the treatment on the distribution can be estimated <u>without any further assumptions</u> (non-parametric estimator; simple treatment-control comparisons)
- We implement this using inverse propensity score weights

IV-QTE:

- Frolich & Melly (forthcoming) develop estimators for unconditional QTE when the treatment (here use of Head Start) is endogenous.
- They extend the LATE approach of Imbens & Angrist (1994) and Abadie (2002) and estimate the effects of HS participation for compliers

Inference:

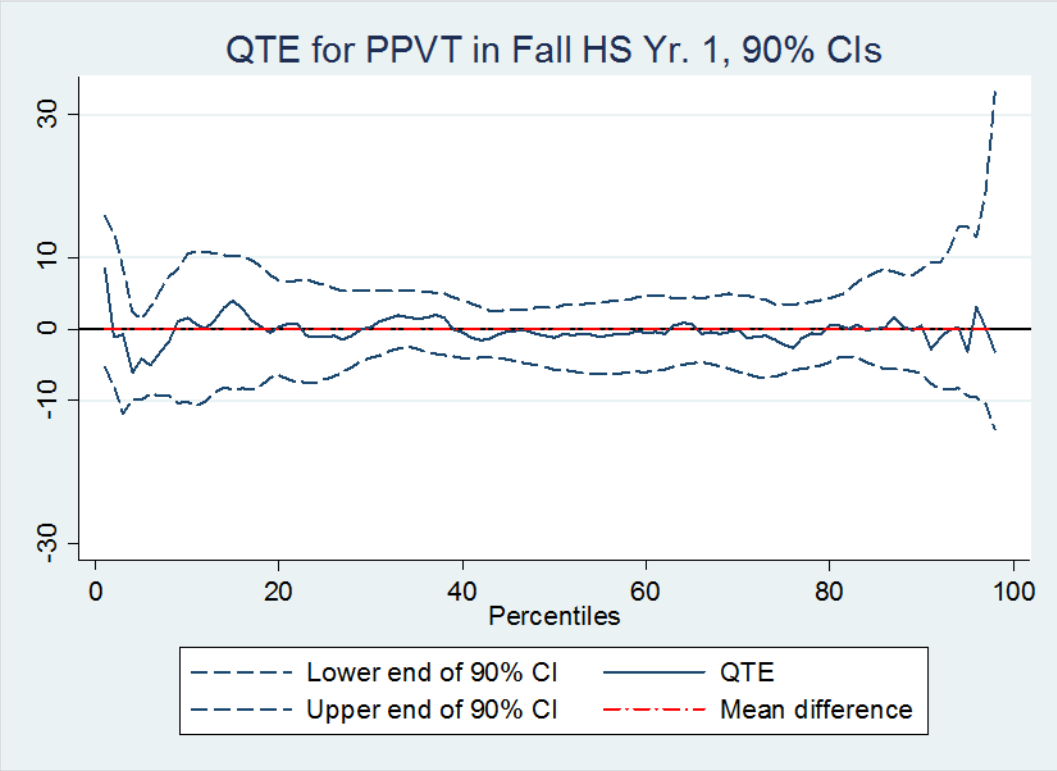
- Bootstrap
- We re-estimate the inverse p-score weights within bootstrap replicates
- Confidence intervals are 90% pointwise



Unconditional QTE for PPVT in end of HS Year

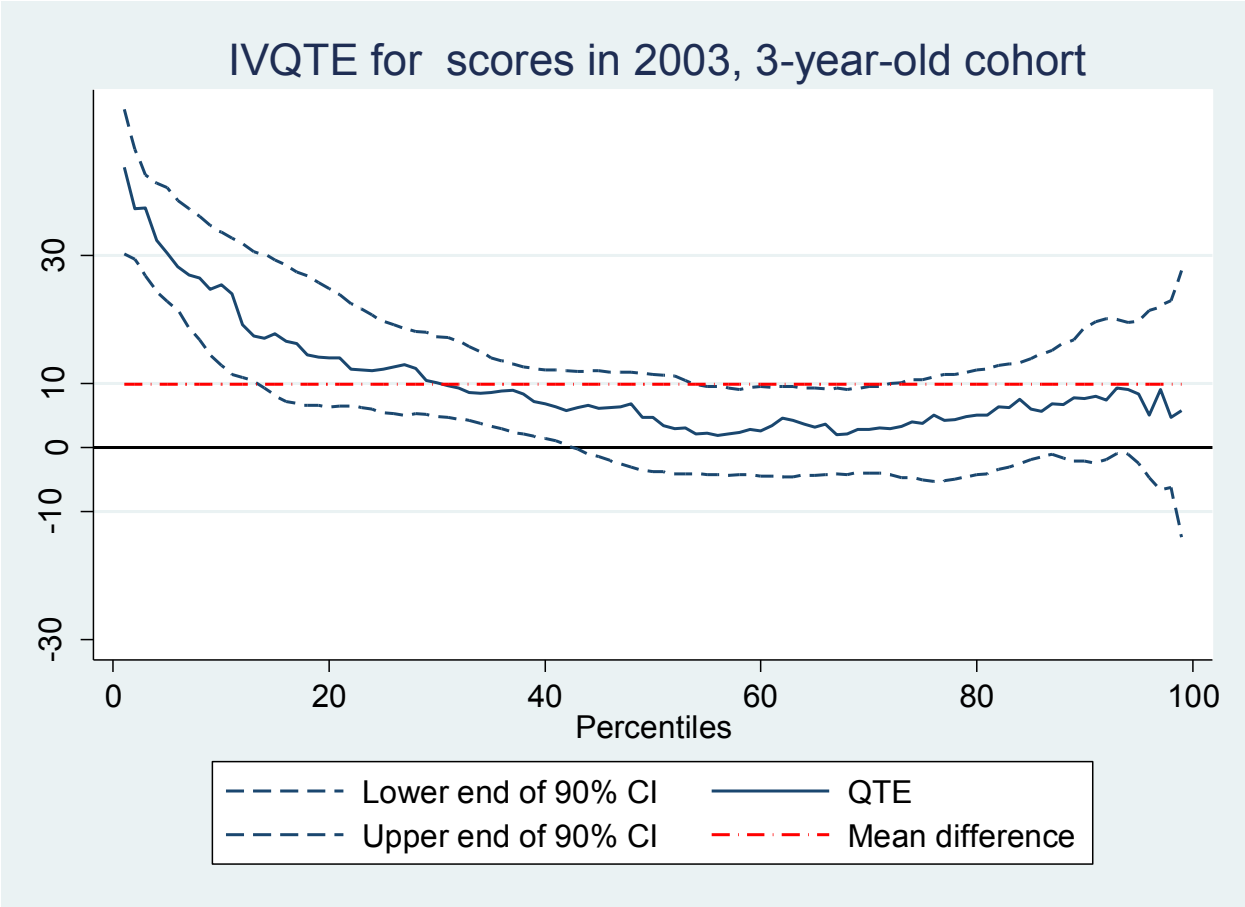
1. Positive cognitive effects across the distribution
2. Gains much larger at the bottom of the distribution
3. Large effects relative to PPVT SD = 38

Balance in baseline scores (Fall 2002)



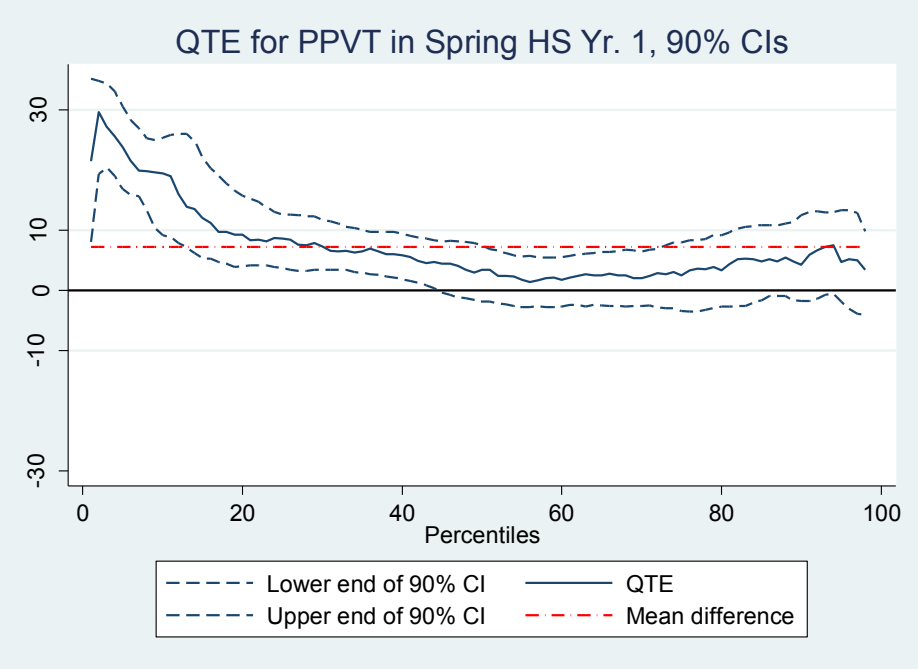
Before moving on, within the distributional approach, we can provide one more balance test

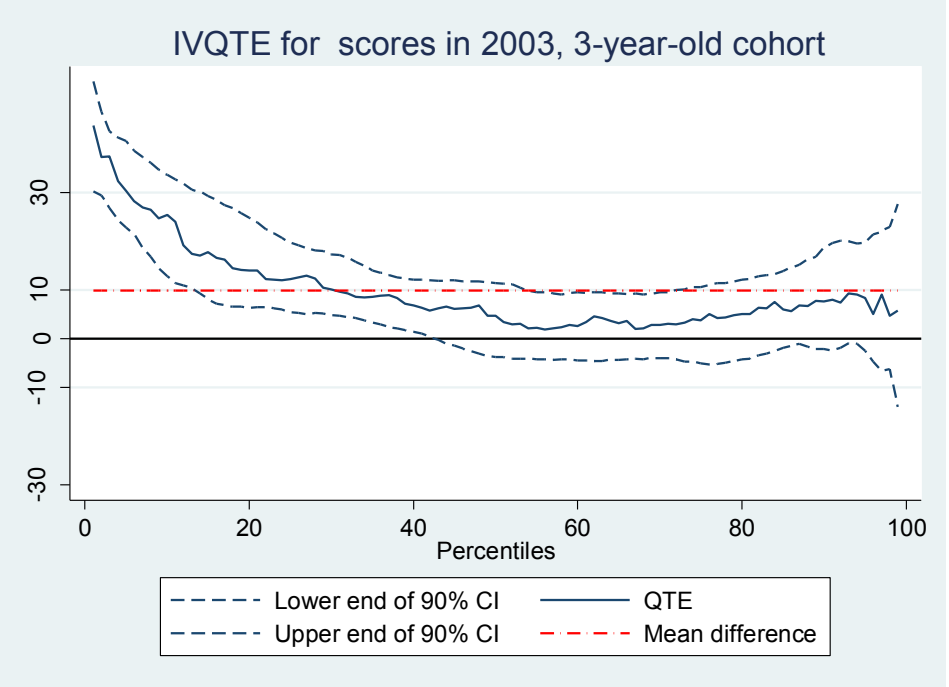
Balance looks great across the distribution for baseline score (estimates are with inverse p-score weighting)



Moving on to IV-QTE:

- Similar to QTE but scaled up
- Large effects: in the bottom quintile from 0.32 to 1 SD
- Clear evidence in favor of the compensatory theory



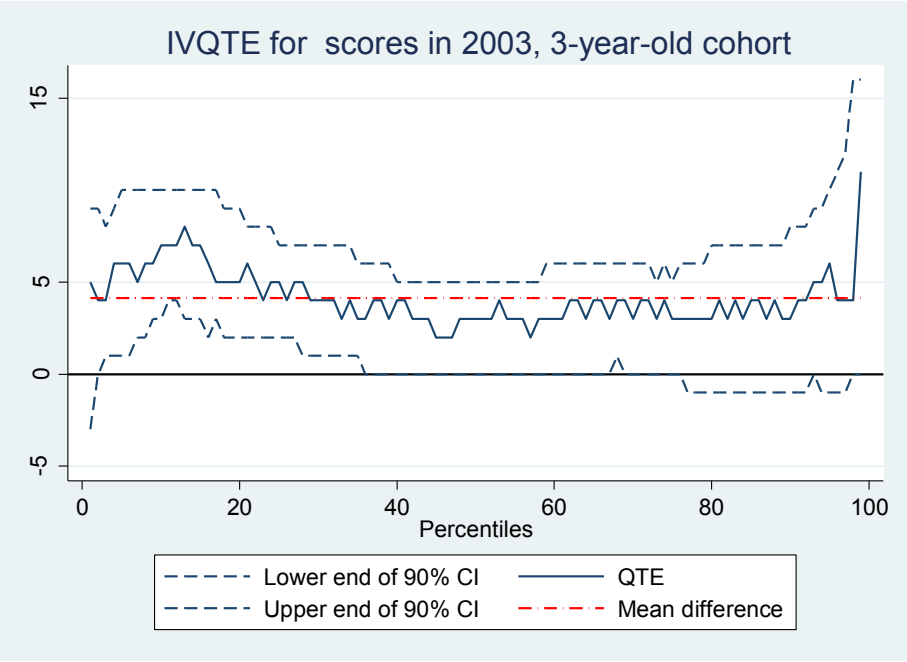


The shape of the QTE and IV-QTE are very similar

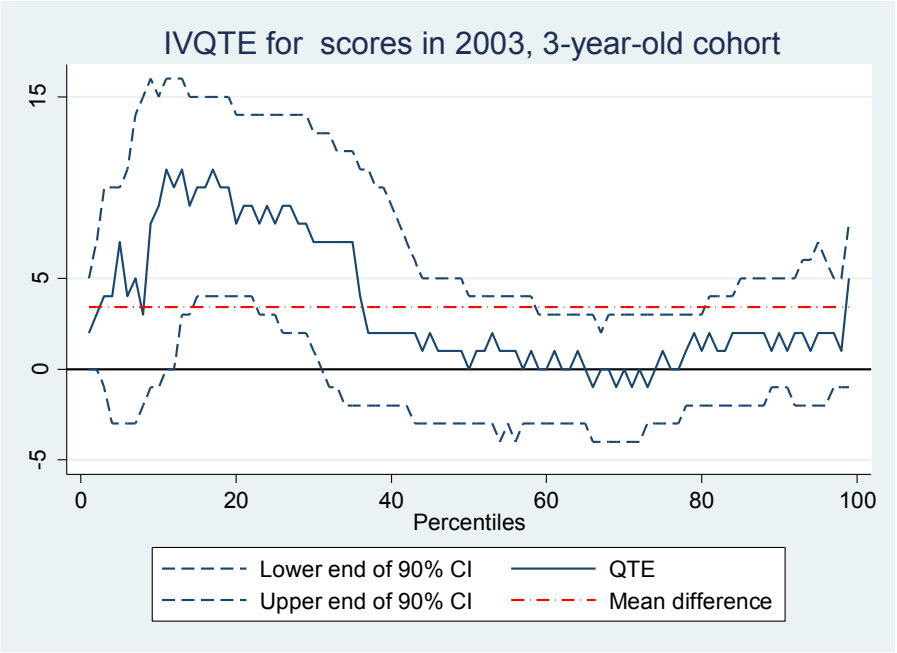
Mostly amounts to a rescaling of the QTE

Other cognitive results for Head Start year

(a) WJIII Pre-Academic Skills

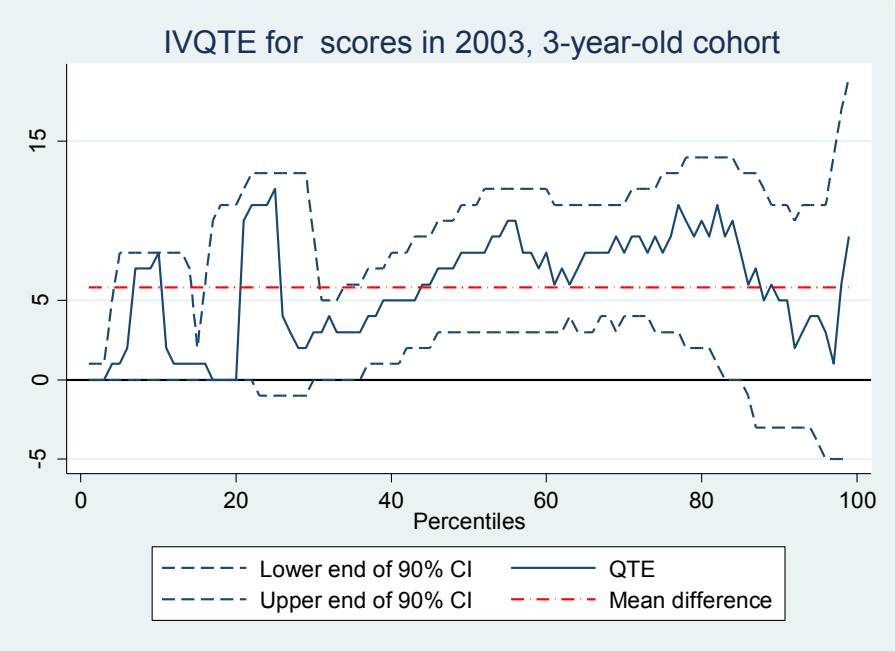


(b) WJIII Applied Problems

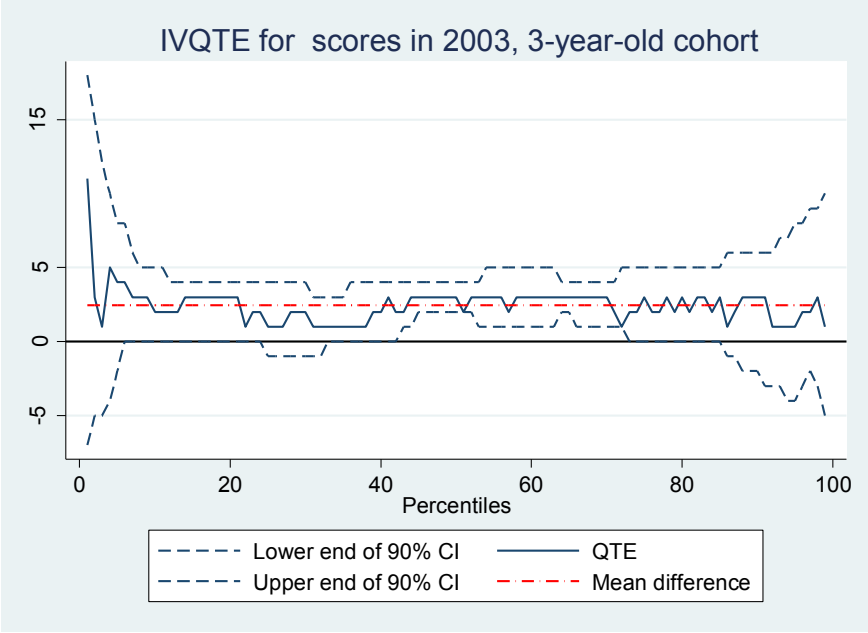


- Woodcock Johnson measures "achievement" including early language (left) and early numeracy (right).
- Results for numeracy are important given Duncan et al's (2007) finding that early numeracy is more predictive of long run achievement
- Overall, similar findings (although less precise) to PPVT

(c) WJIII Letter Word



(b) WJIII Spelling



- Results for "letter-word" show some evidence of larger gains at the top of the distribution

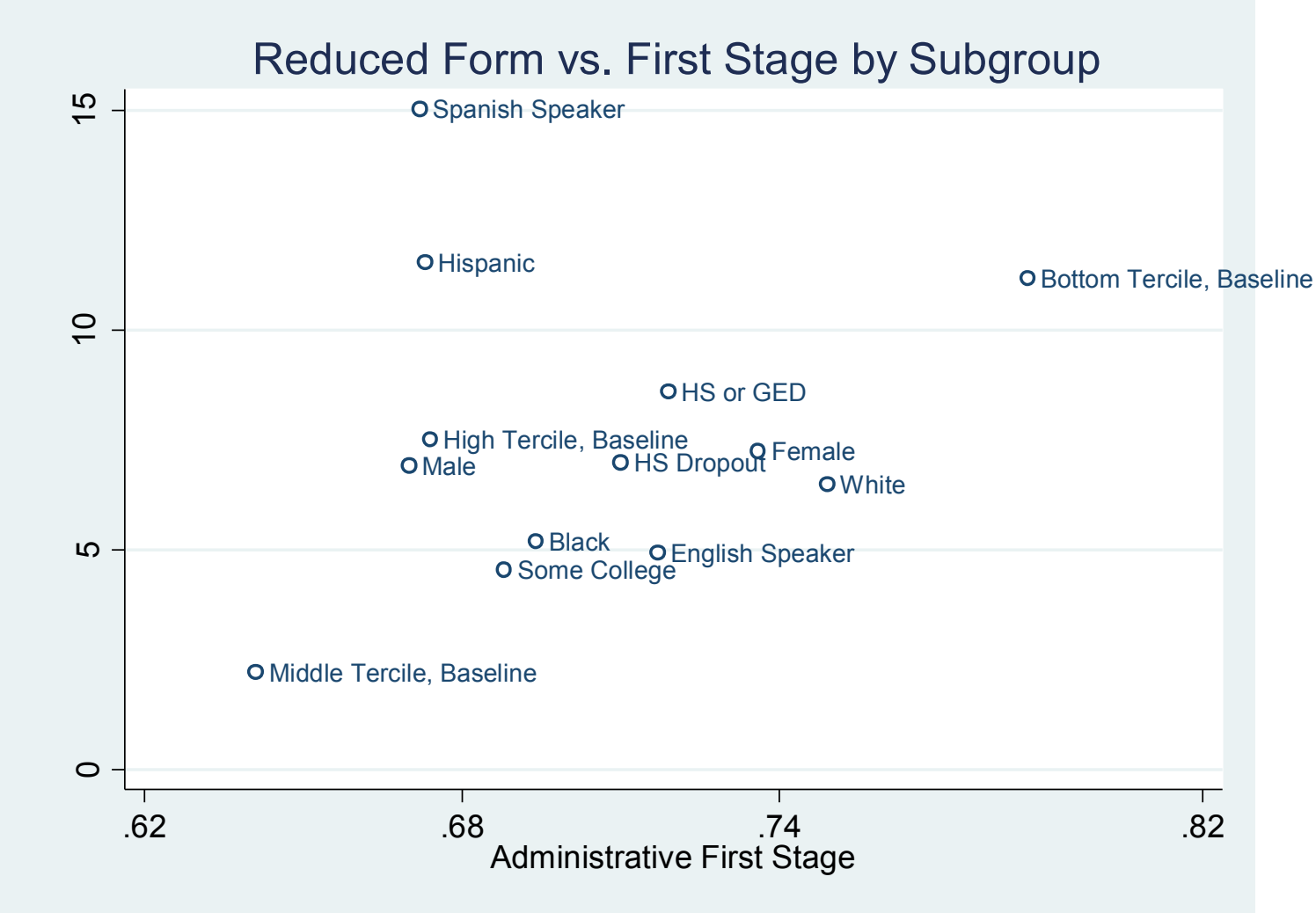
RESULT 2:

Effect of HSIS on cognitive scores during the Head Start year for subgroups

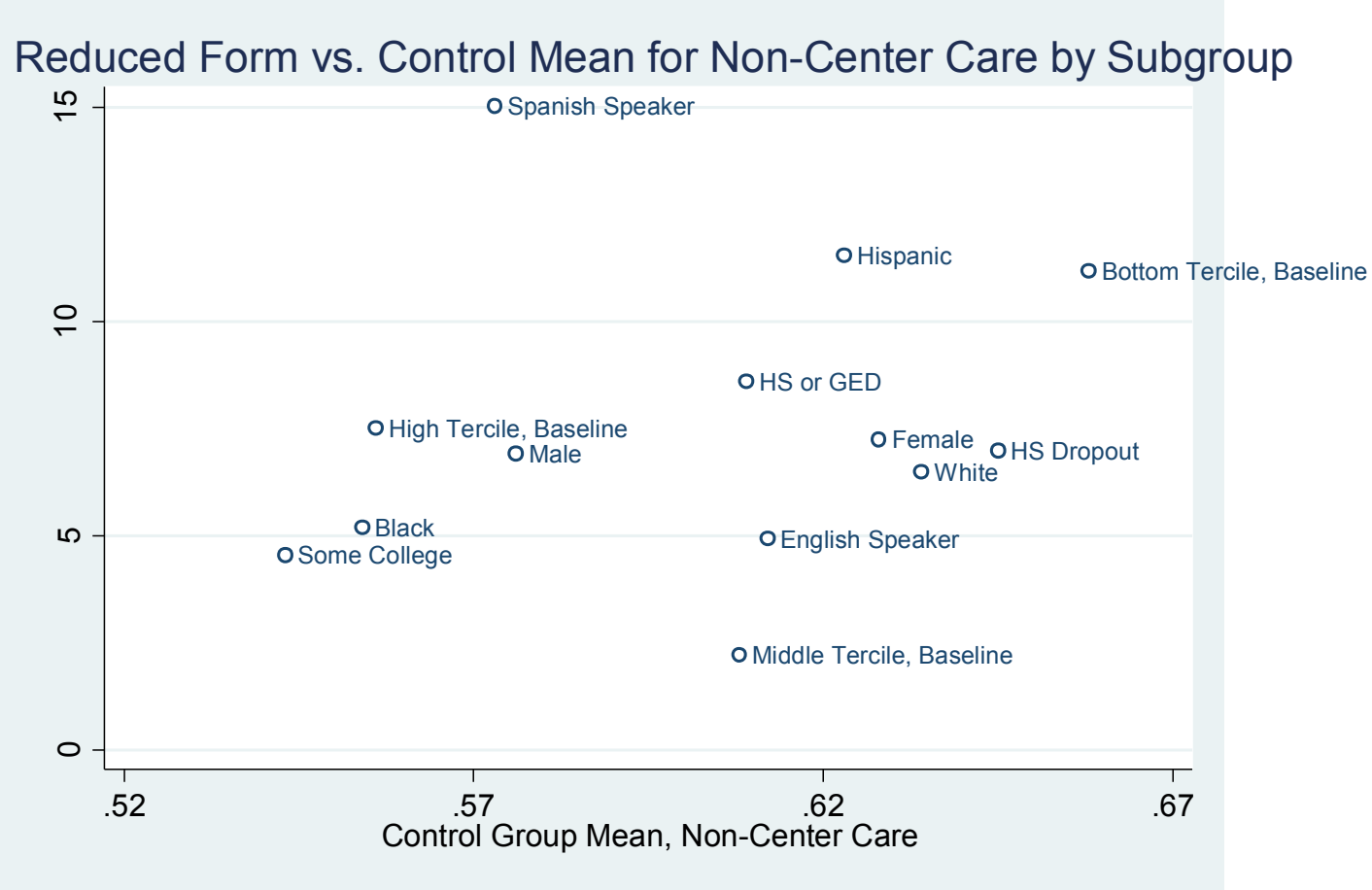
- Another dimension of heterogeneity is to examine effects across groups
- We concentrate here on: race, ethnicity, language and terciles of the baseline score

	Control mean [SD]	Reduced form (SE)	First stage (SE)	Two-stage least (SE)
Race/ethnicity subgroups				
Hispanic	234	11.54***	0.673***	17.13***
	[39]	(3.06)	(0.060)	(4.94)
Non-Hispanic Black	250	5.19**	0.694***	7.48*
	[32]	(2.58)	(0.046)	(3.83)
Non-Hispanic White	268	6.49**	0.749***	8.66**
	[34]	(3.14)	(0.033)	(4.23)
Language subgroups				
Spanish speaker at home	223	15.0***	0.672***	22.37***
	[32]	(3.30)	(0.057)	(5.65)
English speaker at home	261	4.93***	0.717***	6.87***
	[34]	(1.86)	(0.029)	(2.62)
Baseline PPVT score tercile subgroups				
Bottom tercile, baseline PPVT	229	11.2***	0.787***	14.2***
	[33]	(2.77)	(0.035)	(3.73)
Middle tercile, baseline PPVT	251	2.21	0.641***	3.44
	[30]	(2.45)	(0.047)	(3.84)
Top tercile, baseline PPVT	274	7.50**	0.674***	11.13**
	[36]	(3.16)	(0.041)	(4.78)

- Mean effects vary by subgroup – larger for Hispanics [0.3 SD], Spanish speakers [0.4 SD] and bottom tercile [0.3 SD]



- Differing treatment effects across groups do not appear to be driven by differences in the first stage (ITT on y-axis)



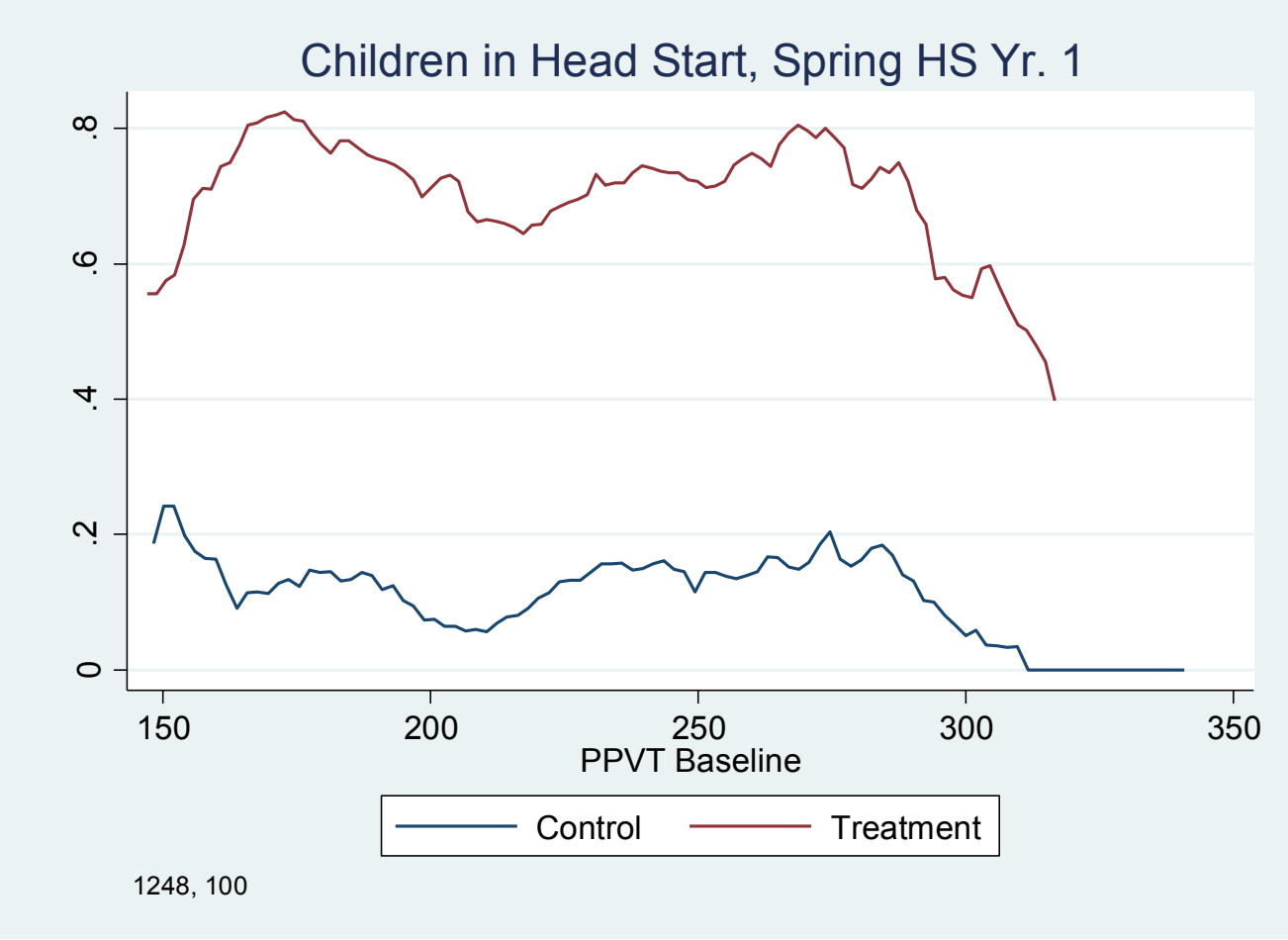
- No evidence that differences are due to the counterfactual care setting (perhaps surprising)

	Control mean [SD]	Reduced form (SE)	First stage (SE)	Two-stage least (SE)
Race/ethnicity subgroups				
Hispanic	234	11.54***	0.673***	17.13***
	[39]	(3.06)	(0.060)	(4.94)
Non-Hispanic Black	250	5.19**	0.694***	7.48*
	[32]	(2.58)	(0.046)	(3.83)
Non-Hispanic White	268	6.49**	0.749***	8.66**
	[34]	(3.14)	(0.033)	(4.23)
Language subgroups				
Spanish speaker at home	223	15.0***	0.672***	22.37***
	[32]	(3.30)	(0.057)	(5.65)
English speaker at home	261	4.93***	0.717***	6.87***
	[34]	(1.86)	(0.029)	(2.62)
Baseline PPVT score tercile subgroups				
Bottom tercile, baseline PPVT	229	11.2***	0.787***	14.2***
	[33]	(2.77)	(0.035)	(3.73)
Middle tercile, baseline PPVT	251	2.21	0.641***	3.44
	[30]	(2.45)	(0.047)	(3.84)
Top tercile, baseline PPVT	274	7.50**	0.674***	11.13**
	[36]	(3.16)	(0.041)	(4.78)

- Significant differences in effect of HS participation across groups
- Additional evidence for compensatory role in Head Start

- The analysis by subgroup shows that those with baseline score in the bottom tercile have a larger first stage (79% compared to full sample 70.5%) and larger ITT (11.2 compared to full sample 7.2). They also have heavy use of non-center care (66% of controls in parent/relative/family day care)
- We explore the differences across baseline scores more nonparametrically using local linear regressions (separately for T and C group)

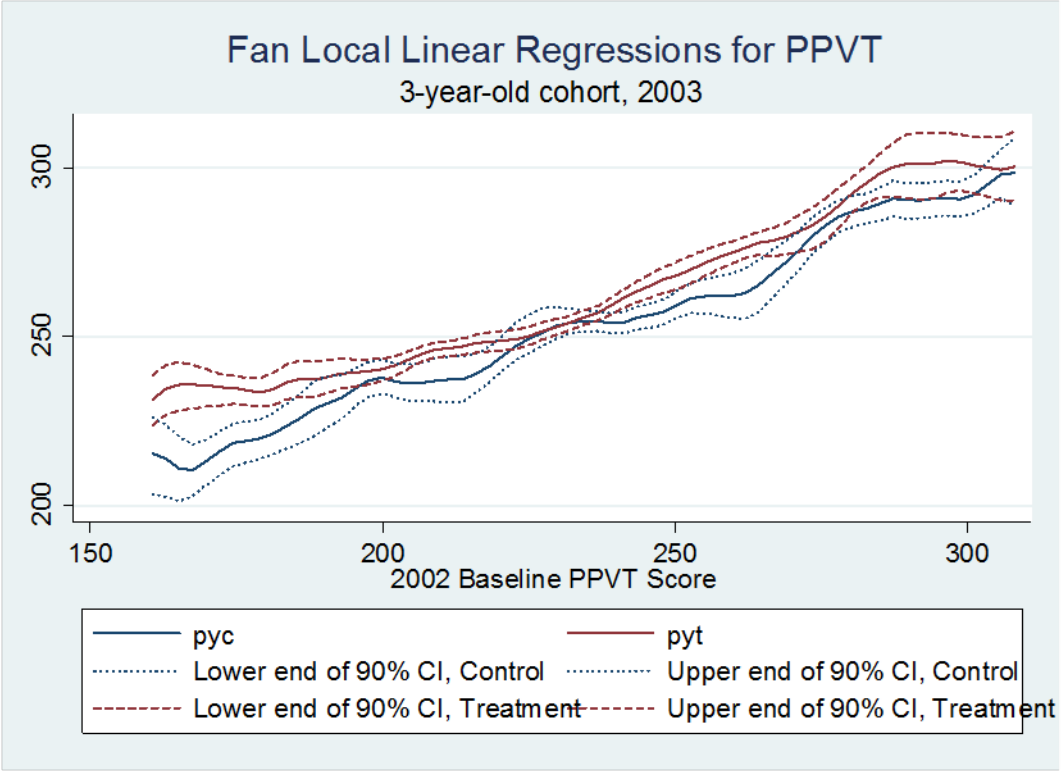
When seen across the entire distribution of baseline skills, the figure shows considerable uniformity in take-up of the offer of HS across baseline skill distribution



Extending to PPVT at the end of HS year

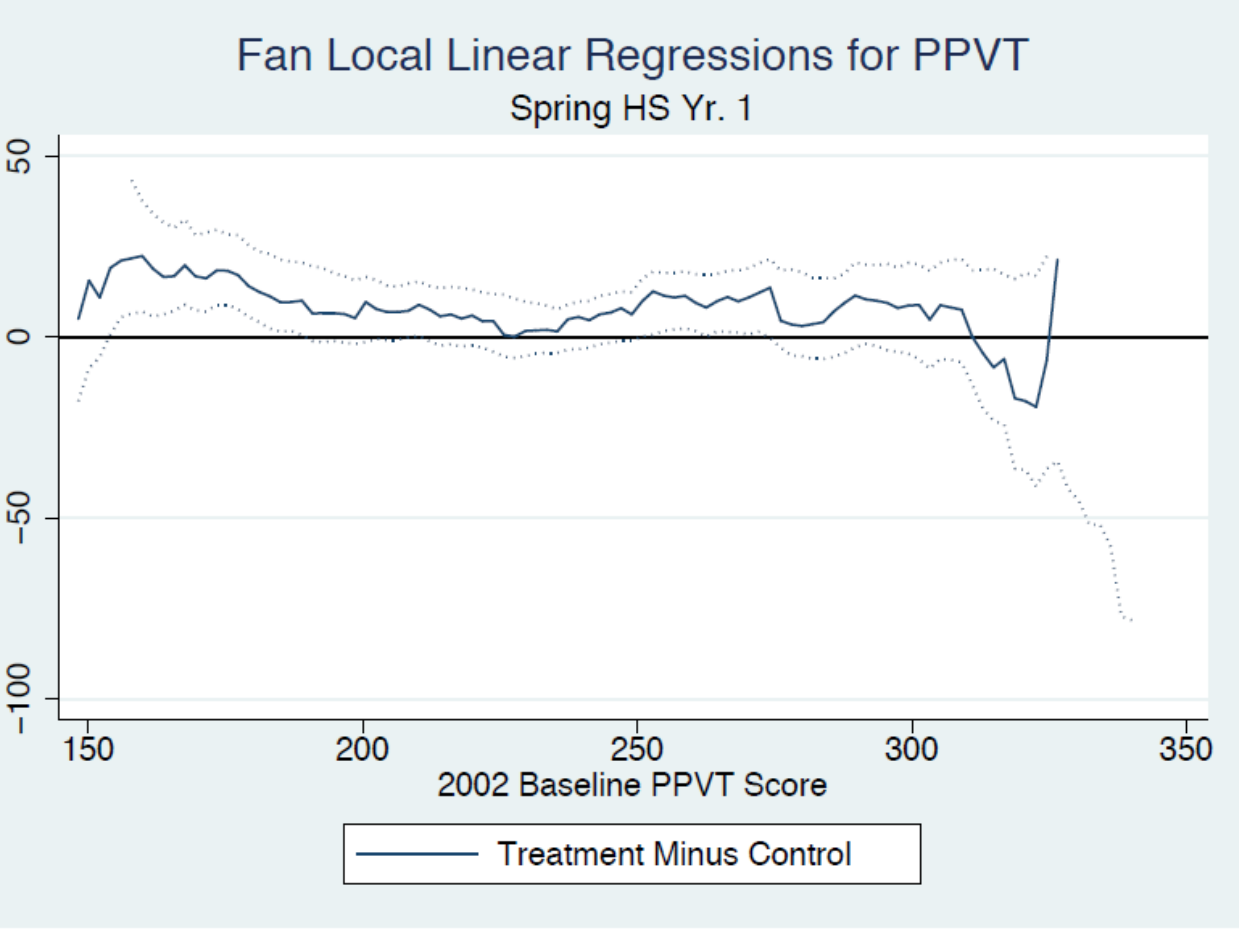
- We estimate the reduced form impacts of a HS offer at the end of the HS year *across* baseline scores (as in Duflo, Dupas and Kremer 2011)

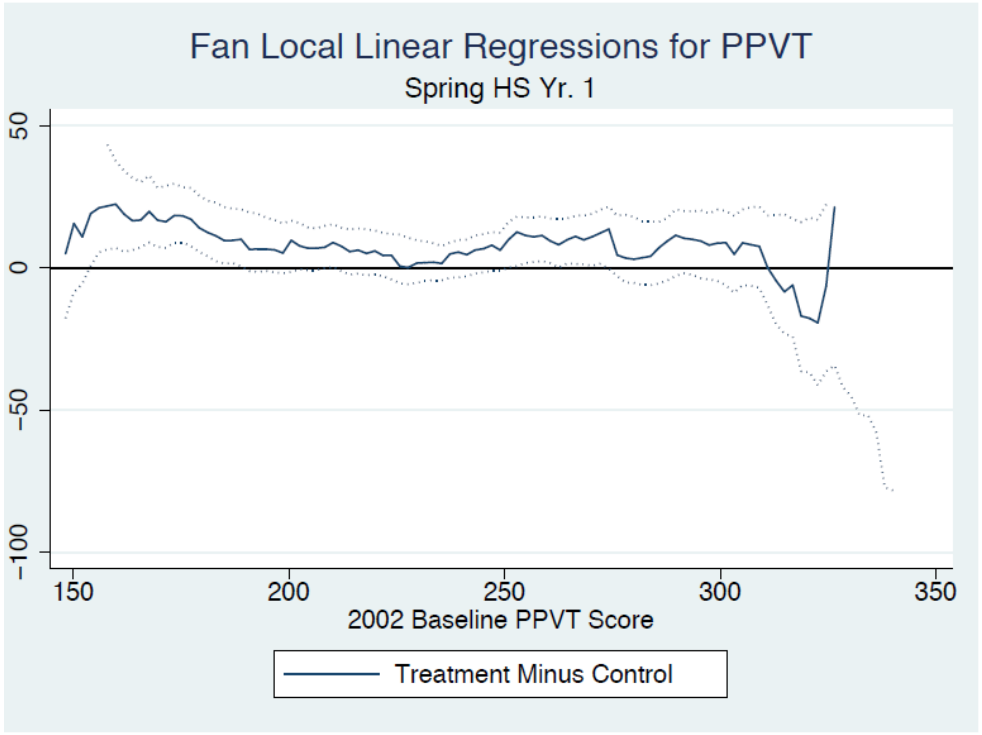
- Estimate local linear regression separately for treatment and control groups
- And then calculate the difference between T and C for a given baseline score.



Local linear regression as alternative

- Here we plot the treatment-minus-control differences
- Again, we find larger effects at the bottom of the distribution (this should be compared to ITT QTE)



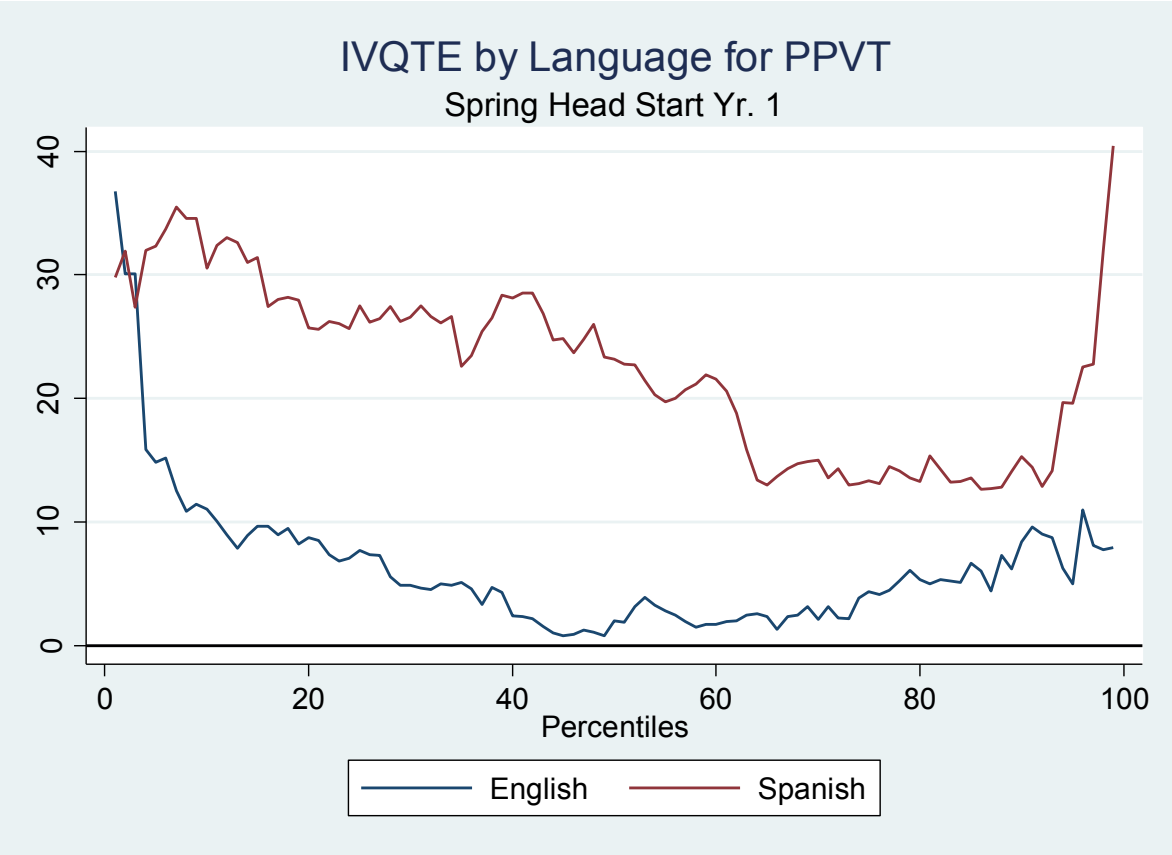


We prefer the QTE approach due to:

- The prevalence of missing PPVT at baseline
- The imbalance of when the baseline assessments were done (leaking into the treatment period)
- The QTE tells us about ex-post changes in achievement gaps, which is of substantive interest

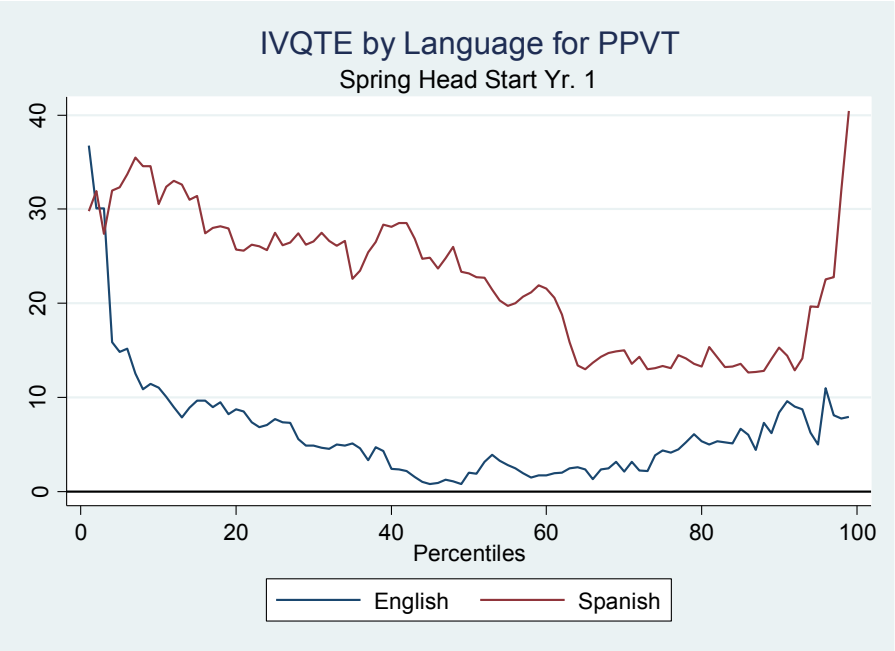
- We can explore further by looking at effects within the demographic subgroups
- Conditional IV-QTEs

IV-QTE by language



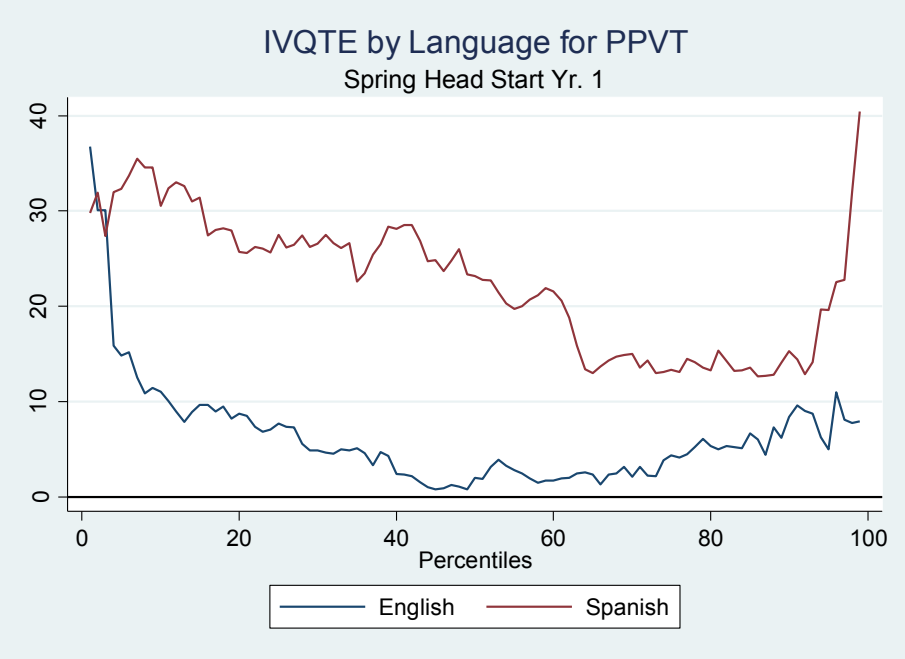
- Dramatic differences across groups: much larger effects for Spanish speakers and further up the distribution → more than 0.3 SD through the 60th percentile

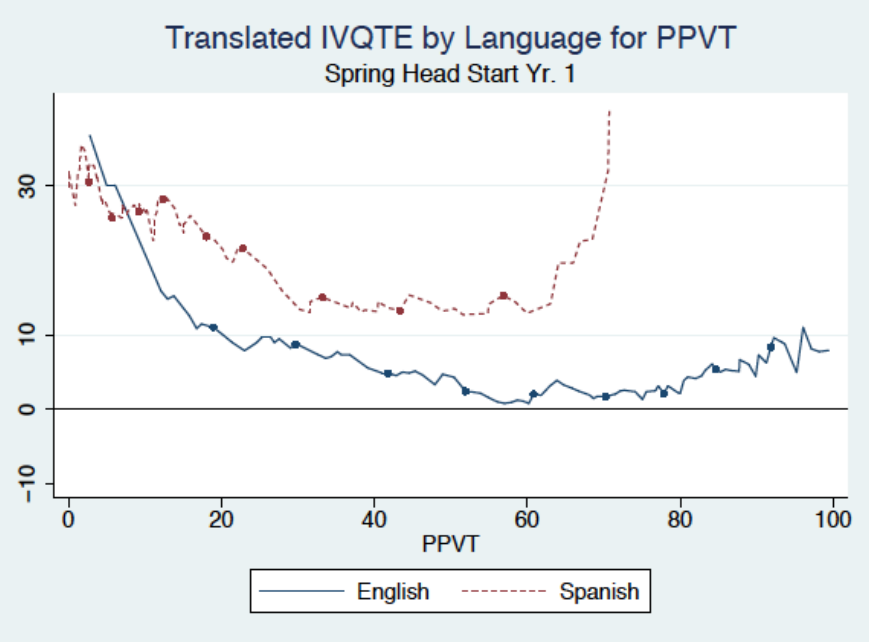
IV-QTE by language



- Examining conditional QTEs is useful
- But one has to be cautious in looking <u>across groups</u> because each is plotted on their <u>own percentile scale</u>
- We also create "translated" QTE where we convert them both to the same absolute scale

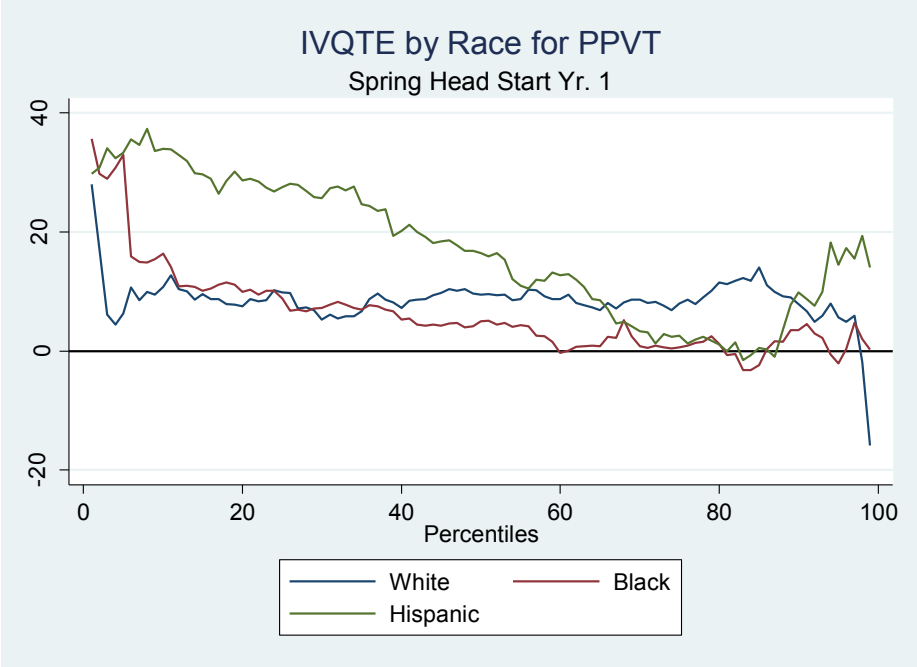
IV-QTE by language

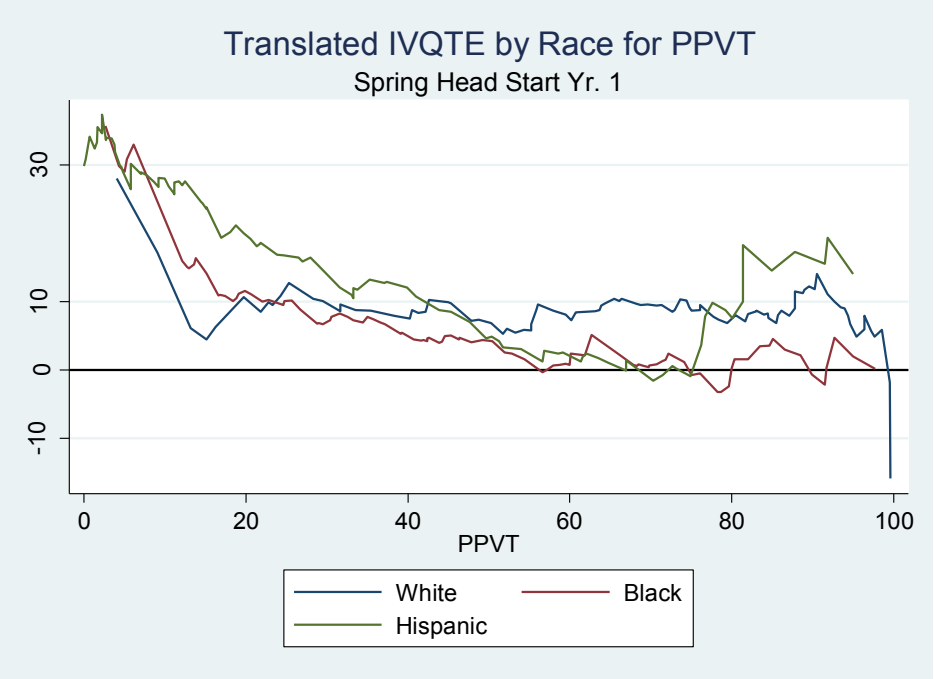




- Translated QTE: same absolute scale is percentiles of the full sample control group
- Translated show same magnitudes in the bottom decile. Larger gains for Spanish speakers up through the distribution.

IV-QTE by race/ethnicity





- Conditional IV-QTE show larger effects for Hispanics, but when put on the same scale they look much more similar

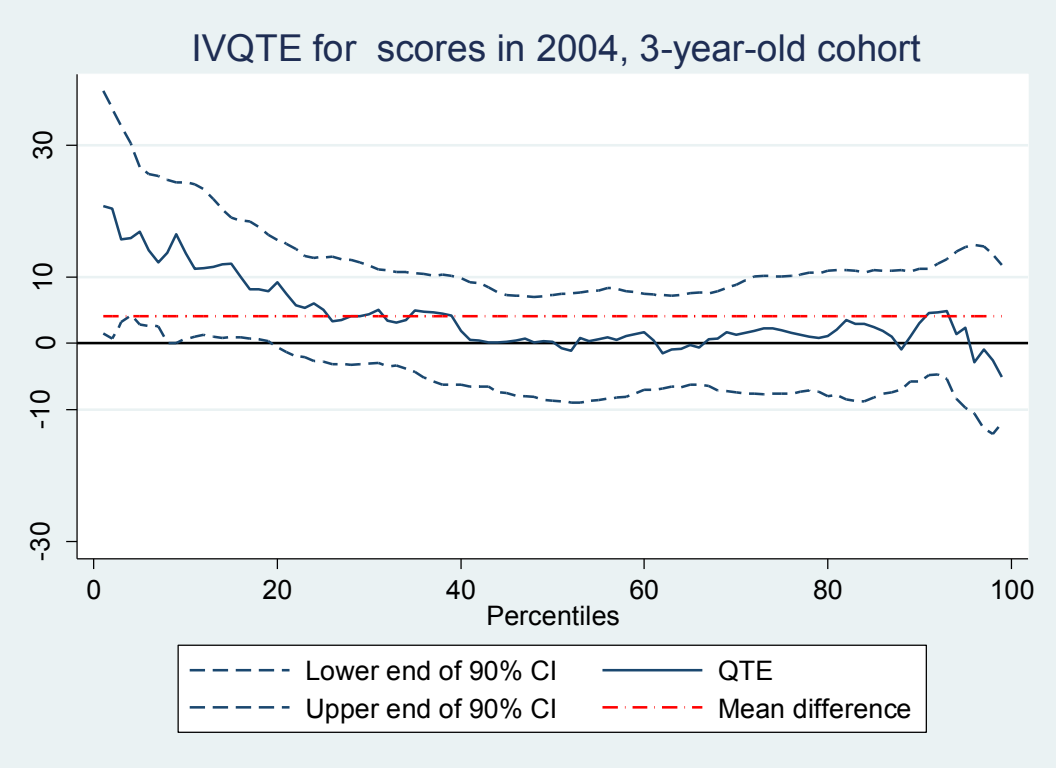
Summary of effects for cognitive outcomes in Head Start year

- HS leads to increases in cognitive skills
- Compelling evidence of heterogeneous effects of HS
- Analysis across groups and across the distribution, we find evidence in favor of compensatory hypothesis
- Gains largest at the bottom of the skill distribution
- Differences in counterfactual care setting do not appear to explain across group treatment effects

RESULT 3:

Effect of HSIS on cognitive scores beyond the Head Start year

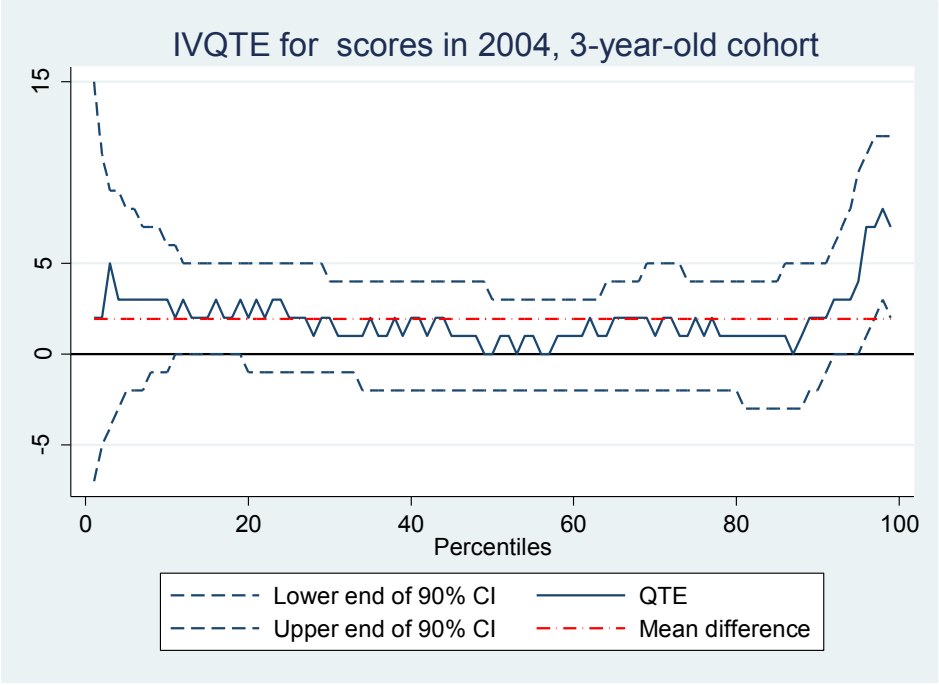
3-year old cohort, Preschool Age 4, PPVT



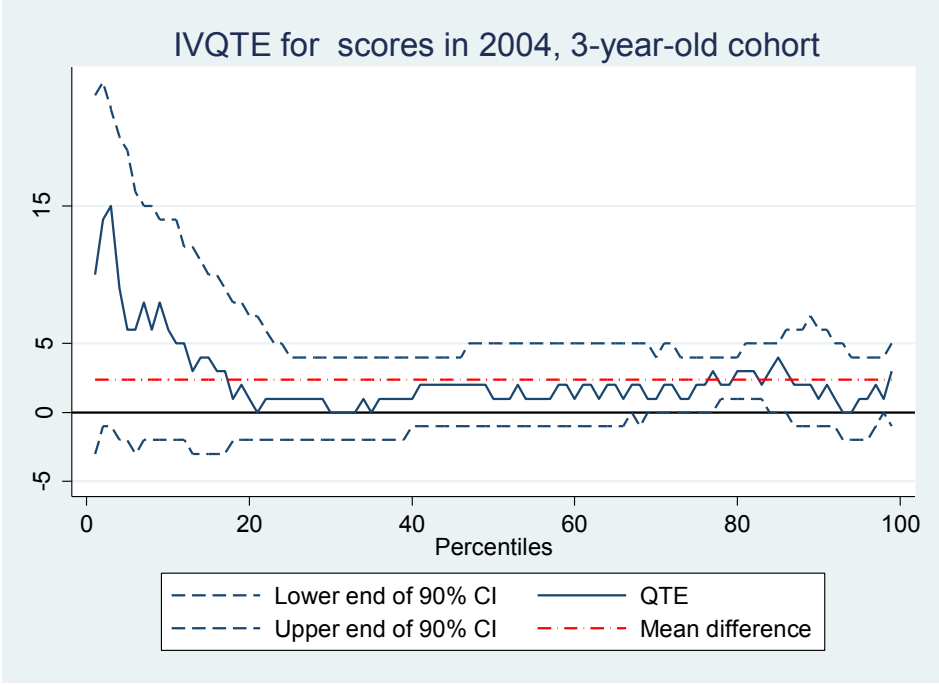
- Positive effects at the bottom persist through the end of preschool years
- [Remember that the first stage here and for all outcomes is HS participation in the 1st year]

3-year old cohort, Preschool Age 4, WJIII

(a) WJIII Pre-Academic Skills

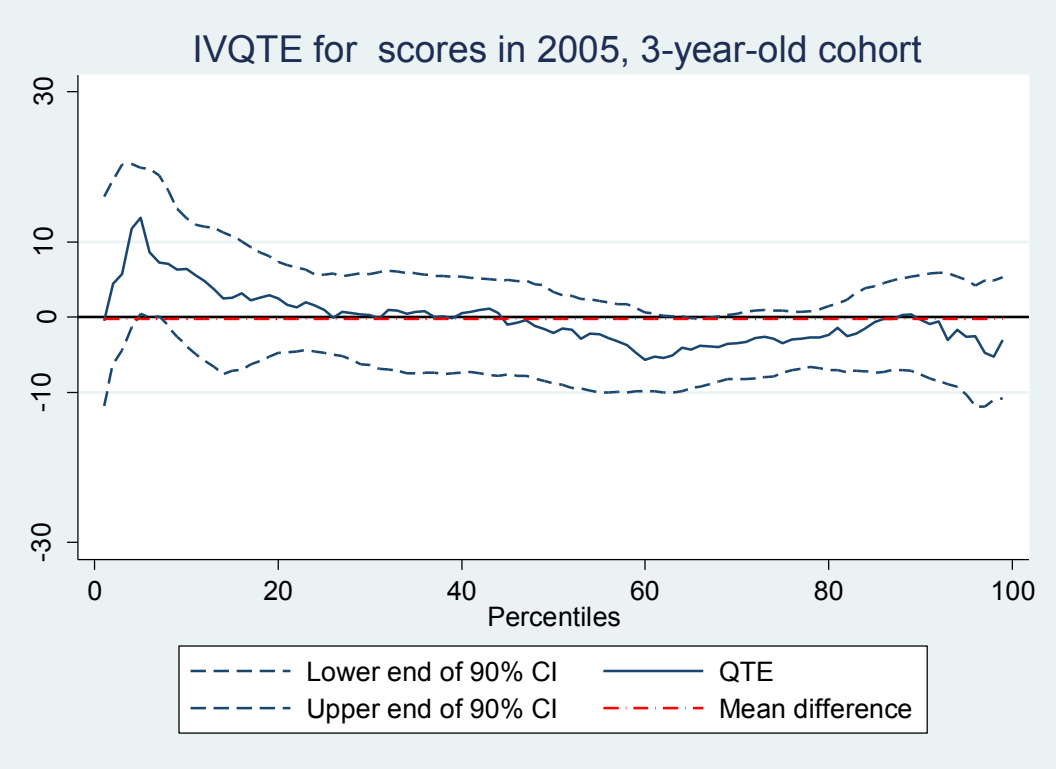


(b) WJIII Applied Problems

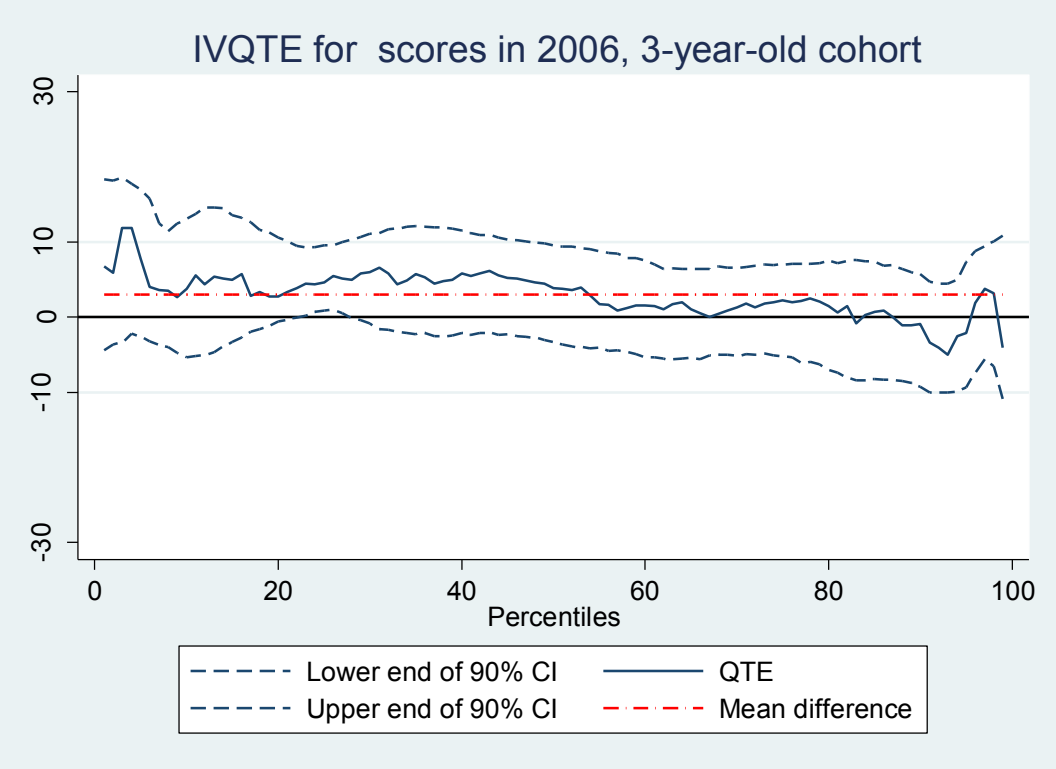


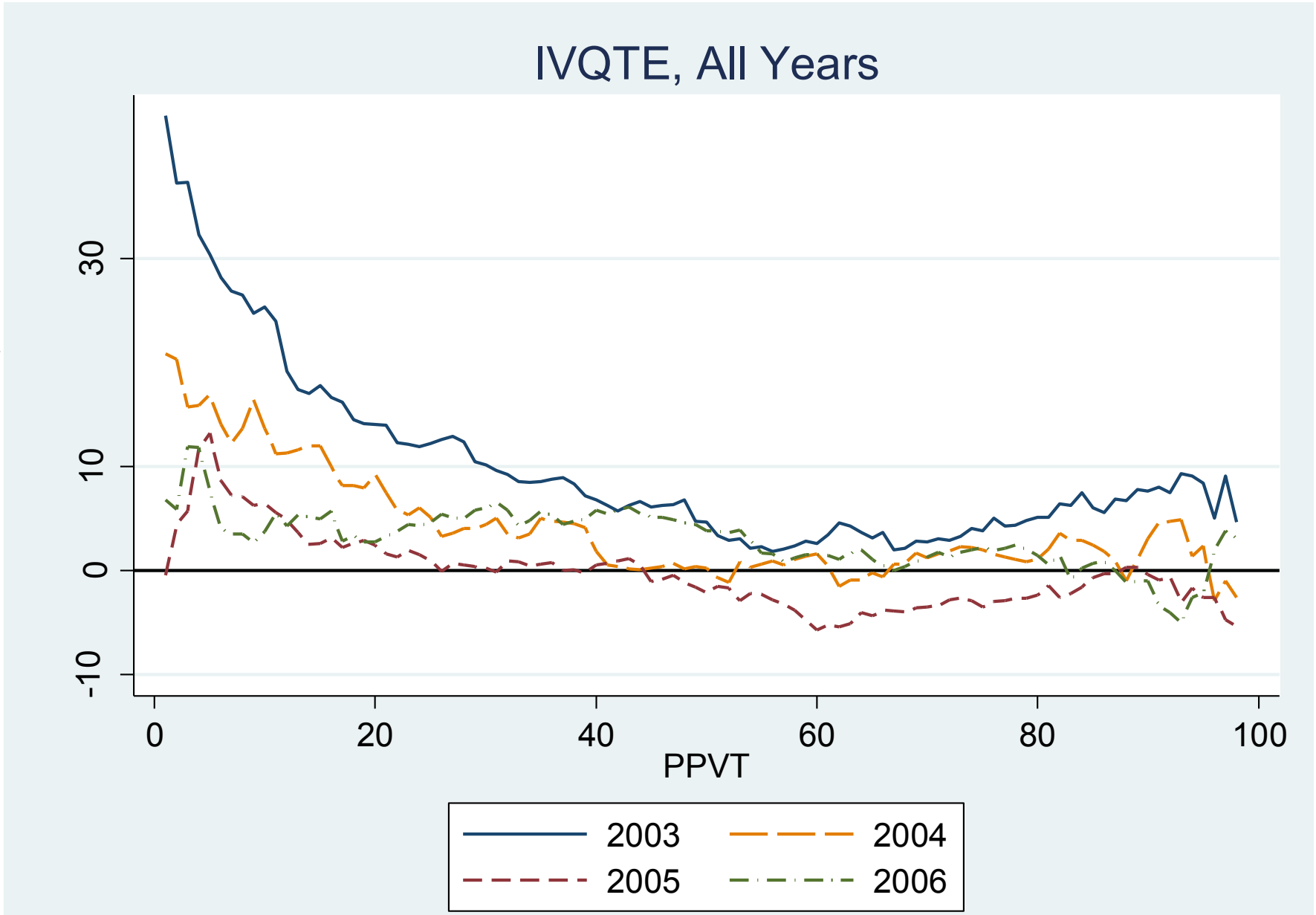
- Same true for WJIII

3-year old cohort, Kindergarten, PPVT

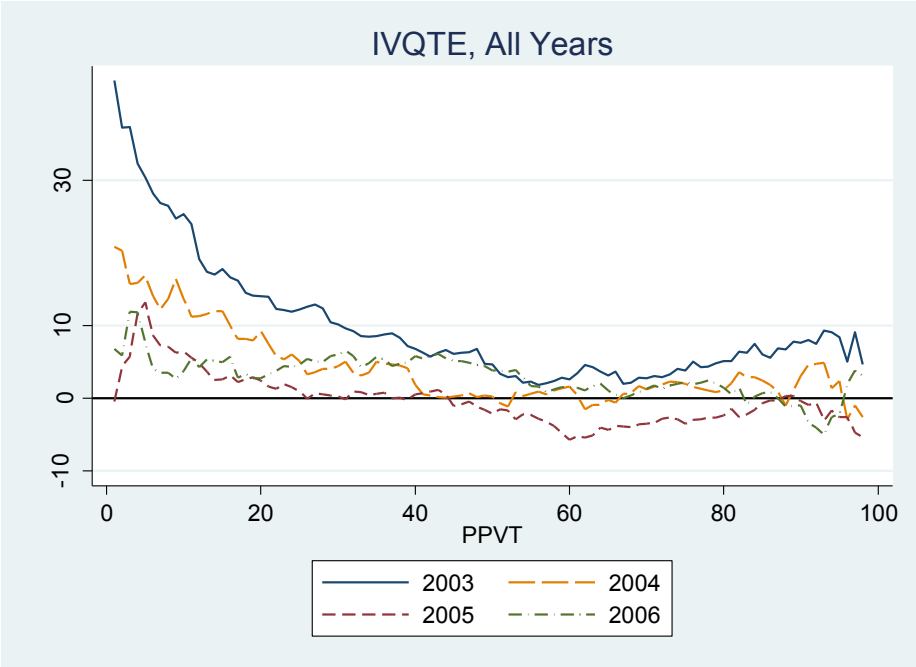


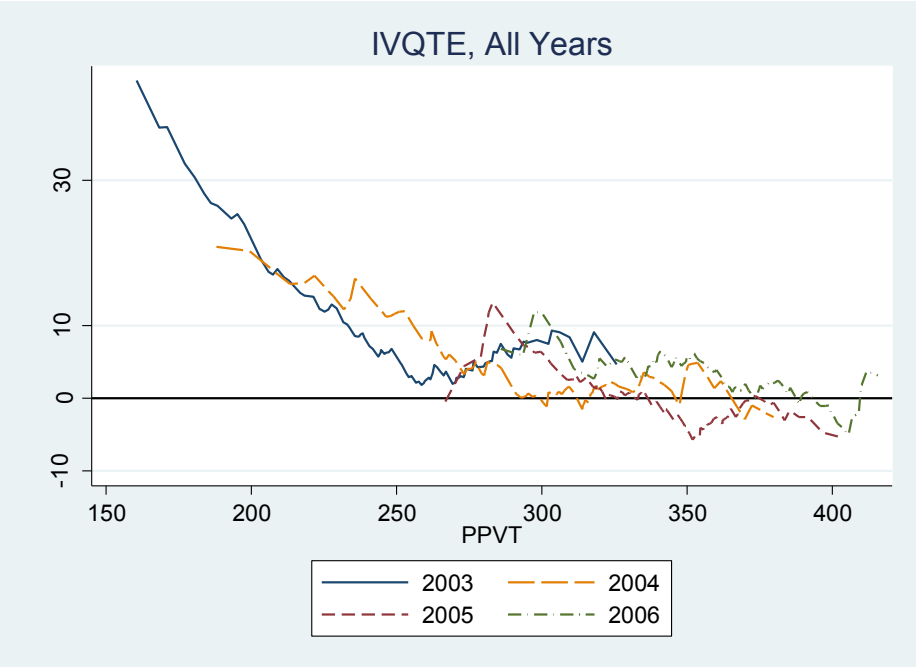
3-year old cohort, 1st grade, PPVT




IVQTE, All Years
30
10
0
-10
0
20
40
60
80
100
PPVT
2003
2004
2005
2006

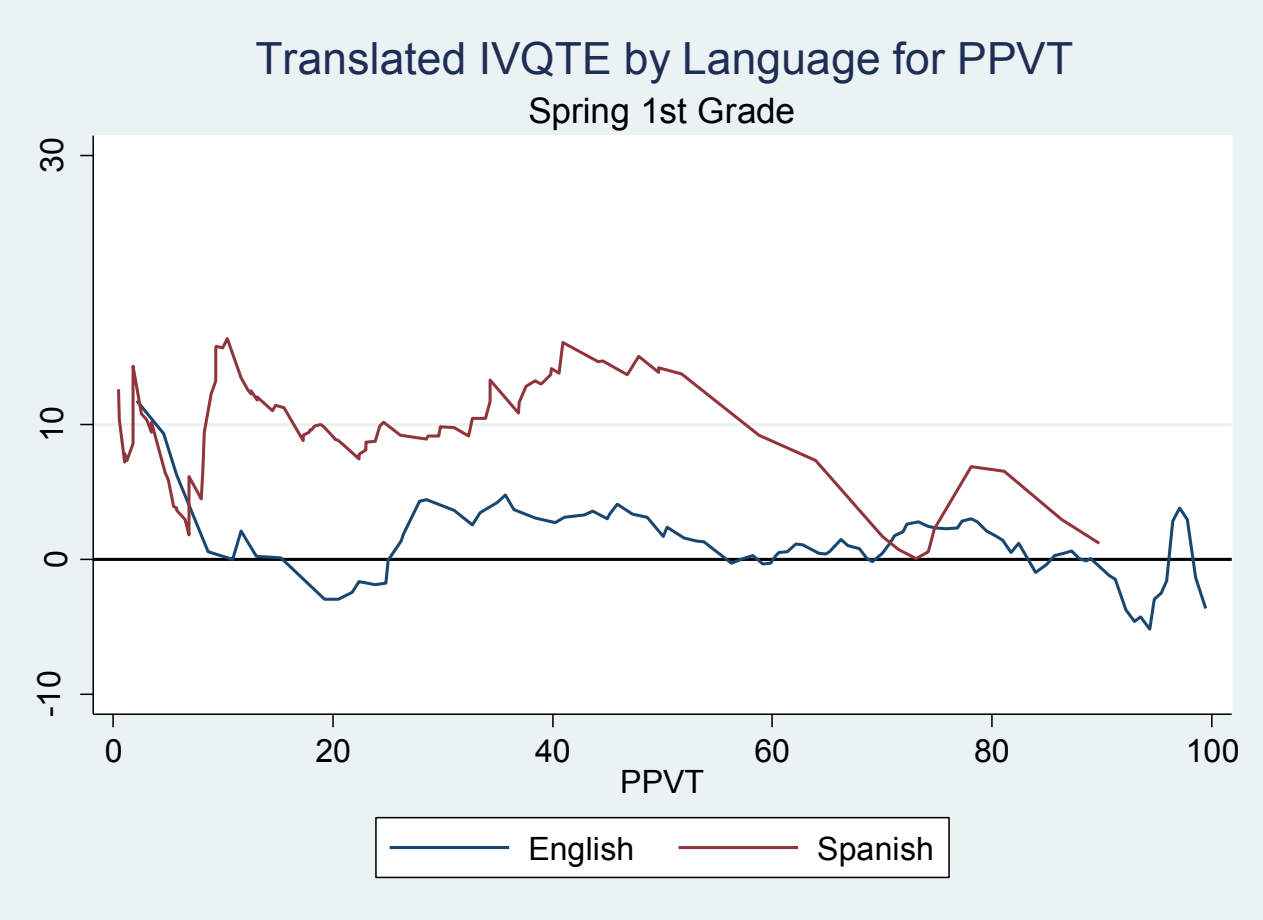
Translated IV-QTE for all years





- Translated graph shows more alignment across years
- The results suggest that HS brings children up to some level but once they are achieving at that level there is no additional gain

Translated IV-QTE by language – 1st Grade



- However, we do find evidence that there is more persistent effects for those with low baseline English language skills (Spanish speakers)

Overall findings for PPVT

- HS participation leads to large effects during preschool years
- The gains are largest at the bottom of the distribution
- Significant fadeout for the full sample but some evidence of persistence for Spanish speakers
  - These persistent effects are large 0.2 to 0.25 SD
- Strong evidence in favor of compensatory hypothesis

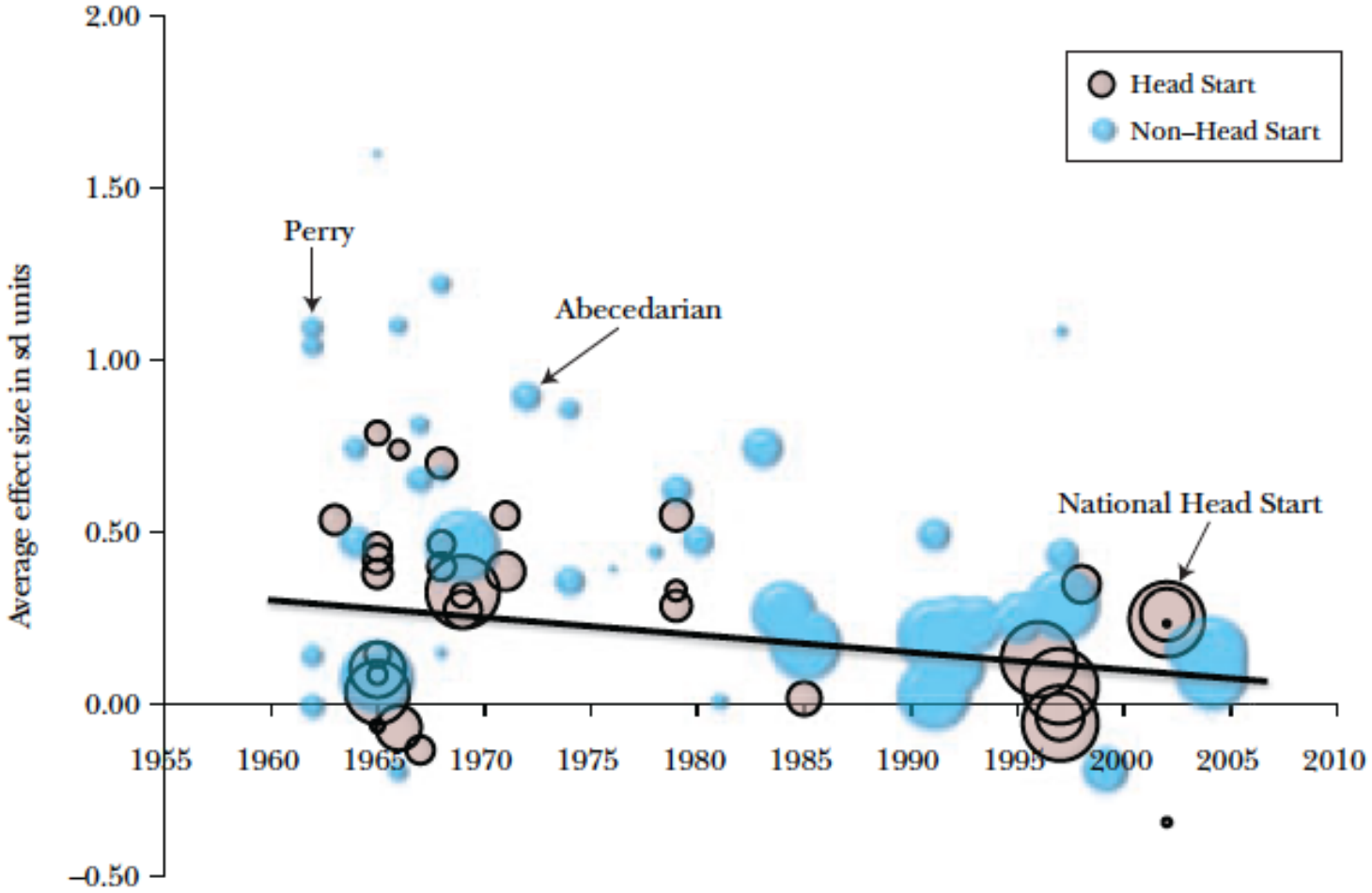
Relating findings to Perry

- Perry found large effects on IQ (0.8 effect size); these are similar to what we found at the bottom of the distribution
  - Notably, Perry targeted very disadvantaged population, IQ 1-2 SD below population mean (Schweinhart & Weikart 1981)
  - Suggest large Perry effects derive from its study population
- Little evidence that the (relatively large) Perry results are due to the difference in the counterfactual care environment.

Figure 2

Average Impact of Early Child Care Programs at End of Treatment

(standard deviation units)



Source: Greg Duncan and Katherine Magnuson "Investing in Preschool Programs," *Journal of Economic Perspectives*, 2013.

What might explain these findings

- Gains from preschool might not persist if the elementary school environment is not of high quality
- OR elementary school teachers are teaching to the lower end of the distribution (and the HS kids are above that point)
- HS teachers may be teaching to some proficiency standard (e.g., knowing ABC, 123) and the quality of the settings are insufficient to achieve gains beyond that point. Cascio and Schanzenbach reference NIEER scores of 5 (on 10 point scale) for HS compared to higher ratings for state funded preschool programs
- Walters (2014) finds some evidence of larger treatment effects in centers with full-day programs and more significant home visiting programs

How do our results speak to the finding of "fadeout and rebound"

- For the full sample, we confirm the findings of fadeout
- BUT we find that looking at mean effects can mask important, lasting, effects across the distribution
  - Hopefully, by following these cohorts we can learn if those with limited English continue to show cognitive gains
- Many have speculated that the long run effects may operate through non-cognitive skills, something we can test using HSIS

2SLS estimates (parent reports)

Table 7: Two-stage least squares estimates of the effect of attending Head Start on socio-emotional outcomes for spring 2003 and 2006

	Head Start year (Spring 2003)	Grade 1 year (Spring 2006)
Parent reports		
Aggressive (ASPI)	-0.115*	-0.084
	(0.068)	(0.078)
Hyperactive (ASPI)	-0.274***	-0.111
	(0.074)	(0.074)
Lack of Social Competencies (ASPI)	0.022	-0.067
	(0.071)	(0.079)
Lack of Social Skills (ASPI)	-0.032	-0.014
	(0.075)	(0.076)
Withdrawn (ASPI)	0.027	-0.069
	(0.071	(0.076)
Conflict (Pianta)	-0.012	-0.169**
	(0.066)	(0.075)
Lack of closeness (Pianta)	-0.121*	-0.099
	(0.064)	(0.076)
Lack of positive relationship (Pianta)	-0.048	-0.166**
	(0.066)	(0.078)
Externalizing behavior problems	-0.169**	-0.106
	(0.070)	(0.072)

We find remarkably little evidence of a treatment effect on non-cognitive skills. [Standardized outcomes, so in SD units.]

2SLS estimates for 1st grade (teacher report, spring 2006)

Teacher reports	
Aggressive (ASPI)	-0.059
	(0.085)
Oppositional (ASPI)	-0.009
	(0.083)
Inattentive (ASPI)	-0.074
	(0.079)
Shy/socially reticent (ASPI)	0.068
	(0.079)
Withdrawn/low energy (ASPI)	0.043
	(0.081)
Problems with structured learning (ASPI)	-0.021
	(0.082)
Problems with interactions (ASPI)	-0.019
	(0.077)
Combined ASPI index—negativity	-0.048
	(0.074)
Combined ASPI index—shy	0.057
	(0.071)
Combined ASPI index—interactive	-0.031
	(0.068)
Lack of closeness (Pianta)	0.019
	(0.087)
Lack of positive relationship (Pianta)	0.016
	(0.089)
Conflict (Pianta)	0.023
	(0.088)

Even less for teacher reports (only available beginning in K)

Conclusions

- We comprehensively examine the effects of Head Start using new data from the Head Start Impact Study, the first large scale randomized evaluation of the program.
- HS participation leads to a large statistically significant increase in literacy and numeracy skills during preschool; the benefits are concentrated in the bottom of the skill distribution and are larger for Hispanics and non-English speaker
- We find no role for differences in take-up or counterfactual care in explaining the results across subgroups
- The cognitive gains fadeout by early elementary school for the full sample but the effects persist for Spanish speakers
- Overall, our evidence is consistent with a compensatory theory of educational interventions with limited evidence for skills-begets-skills