VANDERBILT
PEABODY COLLEGE

PEABODY research institute

Does ECERS Preschool Quality Predict Children's Cognitive Growth: Meta-Analysis of a Dozen Datasets

Dr. Kerry G. Hofer
Vanderbilt University

Dr. Rachel Gordon, Danny Lambouths, and Hillary Rowe
University of Illinois at Chicago

Society for Research on Educational Effectiveness

March 6-8, 2014

Washington, D. C.

Research Team

- Principal Investigators
  - Rachel Gordon
  - Kerry Hofer
- Other Investigators
  - Sandra Wilson
  - Everett Smith
- Graduate Students
  - Danny Lambouths
  - Hillary Rowe
  - Elisabeth Stewart
- Graduate Students (con't).
  - Rowena Crabbe
  - Fang Peng
  - Jenny Kushto-Hoban
  - Ken Fujimoto
  - Anna Colaner
- Consultant
  - Betsy Becker
- Institute for Education Sciences Grant #R305A130118

Broader Project

- Examining ECERS-R and CLASS
- Looking at aspects of validity
  - structural and response process validity
  - predictive validity across a range of outcomes
- Today's focus is on a subset of regression analyses of ECERS-R to cognitive outcomes.
- Policy use of these measures is high-stakes
- Mismatch between their widespread use and the common research conclusions about their potentially weak relationship to children's outcomes

Broader Project

- Using a dozen datasets with focus on replication, including meta-analyses and integrative data analyses.
- Examining linear and non-linear contemporaneous associations, including policy-relevant cutoffs.
- Continued use of the ECERS-R in policy may reflect limitations in the existing literature

Addressing Weaknesses in Knowledge Base

Existing Weaknesses

- Large reliance on single studies
- Lack of policy-relevant conclusions without consideration of cut-scores used
- Frequent combining of quality instruments
- Often only allow for linear relationships
- Focus only on statistical significance

How We Address Concerns

- Leverage a dozen large-scale datasets that span 9 years
- Examination of common and unique outcome measures
- Synthesis of results across original datasets using meta-analytic techniques
- Integrative data analysis, allowing for more power
- Consideration of policy use and practical significance

Questions of Interest

- To what extent do associations of ECERS-R quality to children's cognitive outcomes replicate across studies?
  - Examination of associations and heterogeneity
  - *Approach: Meta-analysis of individual study results.*
- To what extent does evidence support particular policy-relevant cutoffs?
  - Examination of relation to cut-scores and slope differences
  - *Approach: Integrative Data Analysis (Stacked Datasets)*

Word of Caution

- These slides show results of very preliminary investigations. Though we have confidence in what is presented here, these analyses are the first steps towards a more thorough look at the predictive validity of two quality measures. The nuances of the analyses, including the specification of the predictor variable and outcome variables, the details of the regression models, and the meta-analytic techniques used, should all be considered when referring to results presented here.

Variables of Interest

- Predictor
  - ECERS-R (Continuous score)
- Outcome
  - Cognitive achievement
- Covariates
  - Primarily child-level

Datasets

Study Name	Survey Team	Initial Year	Focal Population
3-City Study	RTI	1999	Low-income families from low-income neighborhoods in Boston, Chicago and San Antonio.
ECLS-B	RTI	2001	Nationally-representative sample drawn from birth records in 46 states.
Early Head Start REP	Mathematica	1996-1998	New Early Head Start applicants with a child under 12 months of age.
FACES 1997	Westat	1997	New Head Start 3- and 4- year old participants.
FACES 2000	Westat	2000	
FACES 2003	Westat	2003	
FACES 2006	Mathematica	2006	
Head Start Impact Study	Westat	2002	
Fragile Families	Mathematica	1998-2000	Birth records sampled from hospitals in twenty large U.S. cities.
Multistate Study (NCEDL)	NCEDL	2001	State pre-kindergarten programs randomly selected from four states (IL, OK, KY, GA) and from regions in two states (CA, NY)
PCER	IES,Sites,RTI Mathematica	2003	Twelve sites implemented curricula in preschool programs. Each site had 14-20 programs.
QUINCE	FPG	2004	Twenty-four CCR&R agencies in five states (CA,IA,MN,NE,NC)

Datasets (with Outcomes of Interest)

Study Name	PPVT	WJ LW	WJ AP	WJ D	WJ S	BRCK	PLS	ECLSB-READ	ECLSB-MATH
3-City Study		X	X						
ECLS-B								X	X
Early Head Start REP	X	X	X						
FACES 1997	X	X	X	X					
FACES 2000	X	X	X	X					
FACES 2003	X	X	X	X					
FACES 2006	X	X	X		X				
Fragile Families	x								
Head Start Impact Study	X	X	X		X				
Multistate Study (NCEDL)	X		X						
PCER	X	X	X		X				
QUINCE						X	X		

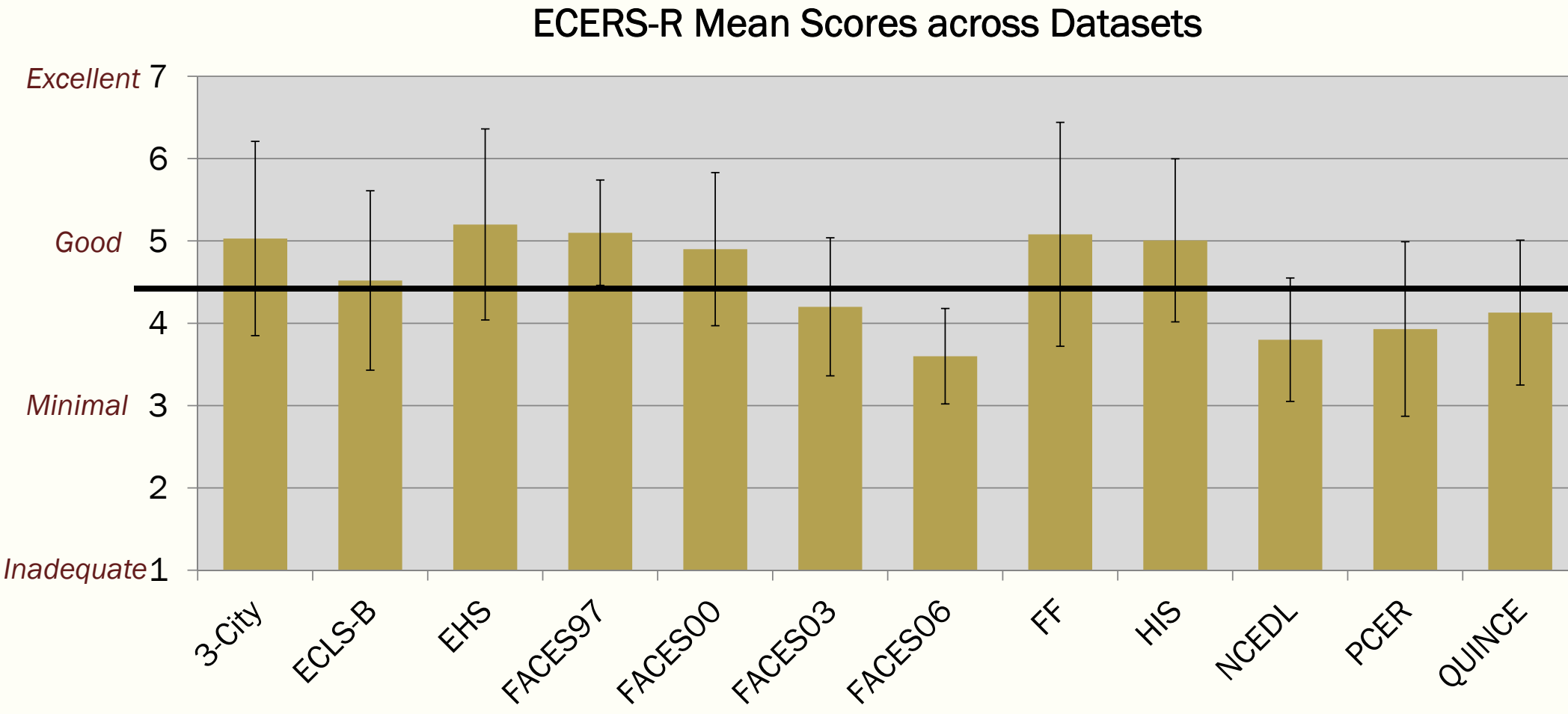
Participants

Child Characteristics	N (Datasets)	MIN	MAX	MEAN	SD
Age at Assessment (Years)	11	2.96	5.11	4.53	.61
% Female	12	.45	.52	.49	.02
% Hispanic	12	.11	.34	.22	.08
% White	12	.09	.57	.31	.13
% Black	12	.16	.61	.36	.13
% Other	12	.01	.17	.08	.05
% Low Income Families	12	.05	.99	.72	.30
% With Disability	10	.01	.26	.14	.08

Descriptives - Outcomes

Primary Outcomes	N (Datasets)	MIN	MAX	MEAN	SD
PPVT	9	86.06	95.37	89.92	3.51
WJ Letter-Word ID	8	90.70	103.27	95.83	4.40
WJ Applied Problems	9	86.75	98.26	91.23	3.85
WJ Dictation	3	86.26	87.65	87.02	0.71
WJ Spelling	3	93.51	96.77	95.37	1.68
Bracken	1	-	-	104.08	-
Preschool Lang.Scale	1	-	-	99.82	-
ECLS-READING	1	-	-	27.14	-
ECLS-MATH	1	-	-	30.88	-

Descriptives – Quality Predictor



Analytic Plan – General Overview

- Using multiple lenses through which to view replication
  - **Step 1:** Bivariate Models (no covariates)
  - **Step 2:** Basic Models (demographic covariates only)
  - **Step 3:** Gain Models (demographics + pretest covariates)
  - **Step 4:** Integrative Analysis Approach

ANALYSIS: STEPS 1-3 (USE OF META-ANALYTIC TECHNIQUES)

Analytic Plan – Steps 1-3

- Steps 1-3 involve the synthesis of information across individual regression models run by dataset/outcome.
- Use of standardized regression coefficients (and associated standardized standard errors) and fixed-effect meta-analyses
- Focus on linear associations in the presentation.
  - Also examined non-linearities in multiple ways, but have found limited evidence for such specifications to date.

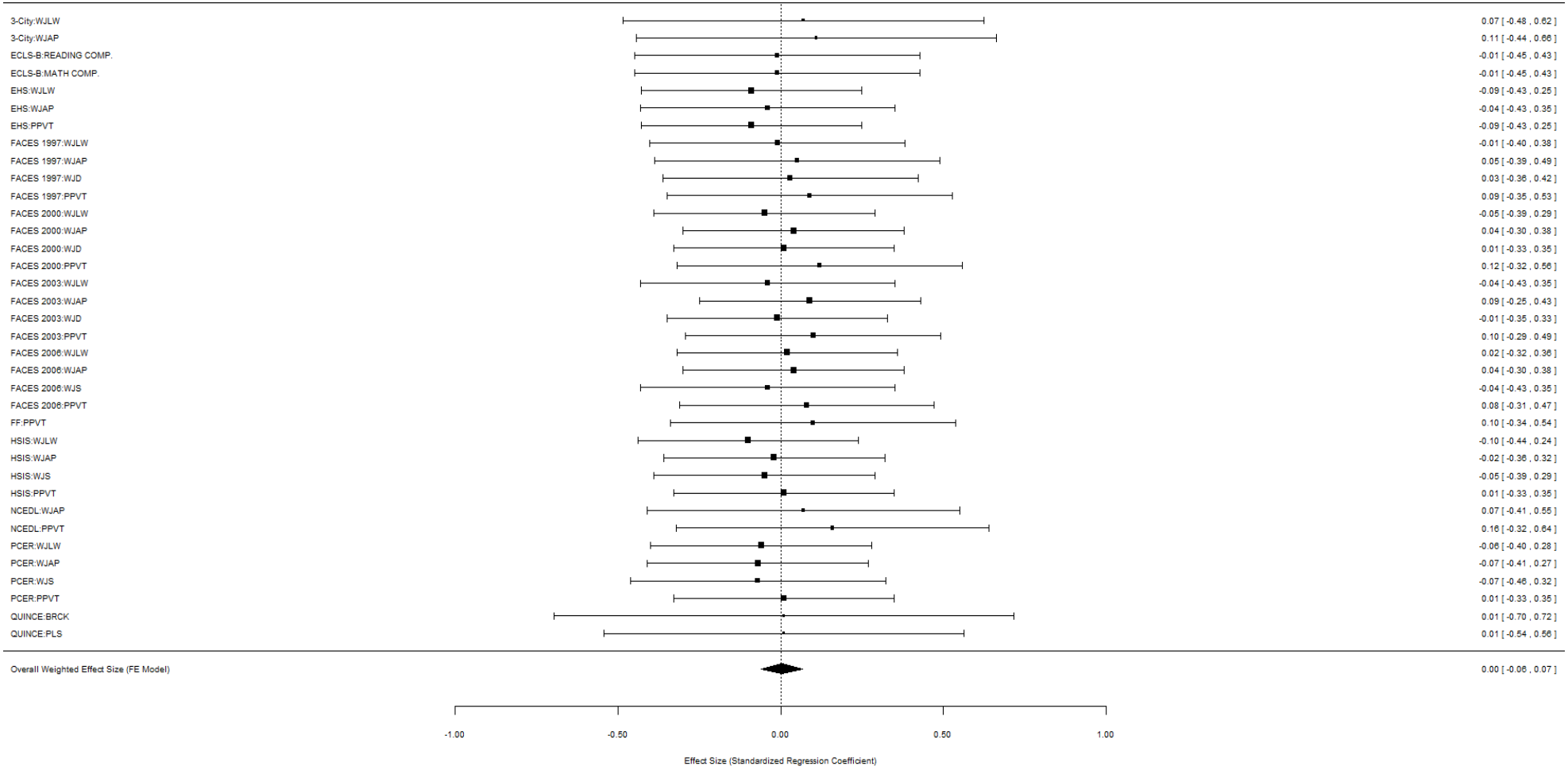
Details: Step 1 (Bivariate Models)

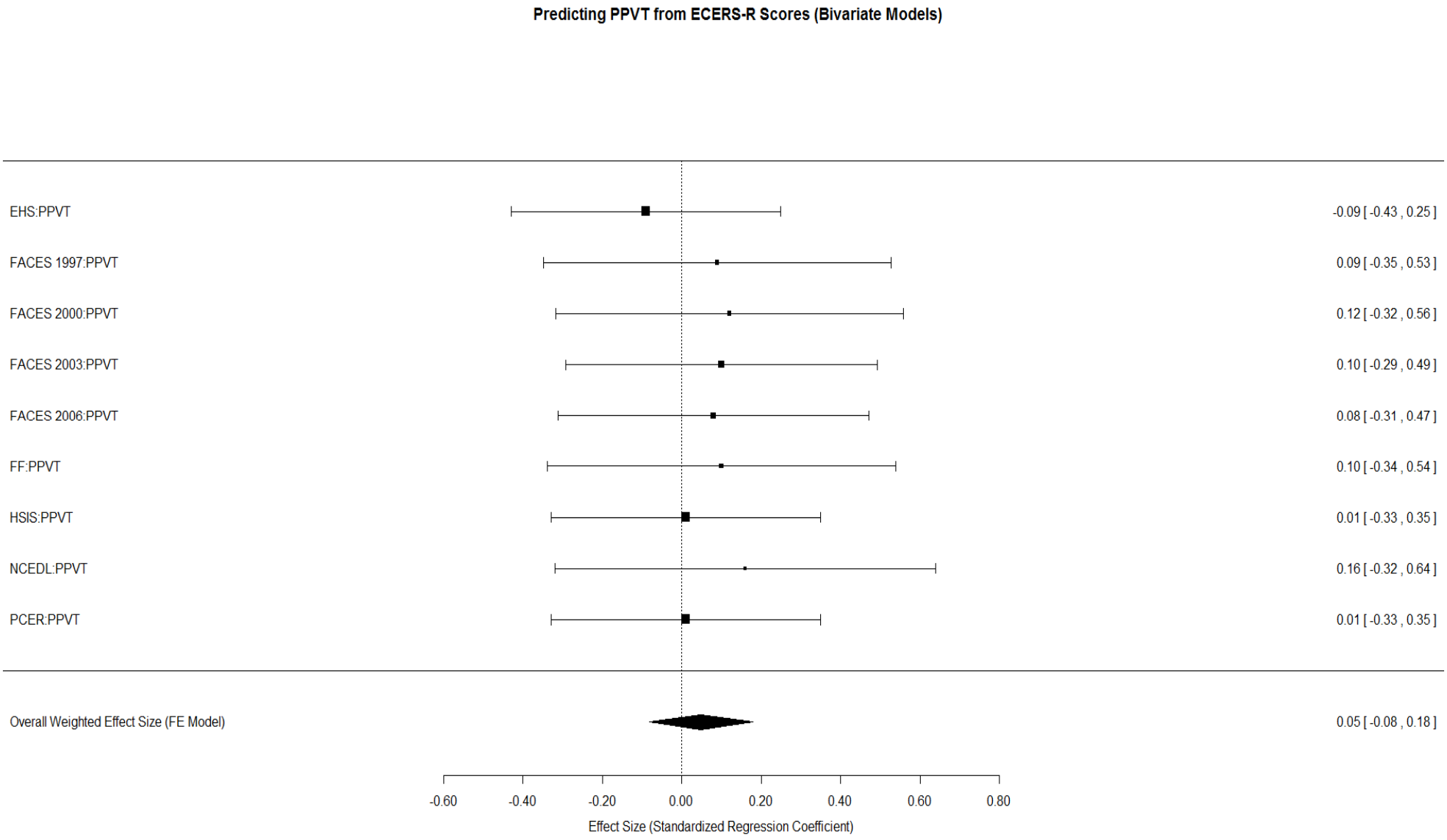
Study Name	Unweighted # of Children	Unweighted # of Classrooms	Multilevel Dataset	Sampling Weight s Used	Robust SE Used
3-City Study	161	-	N	N	N
ECLS-B	1250	-	N	Y	N
Early Head Start REP	816	-	N	N	N
FACES 1997	1133	392	Y	Y	Y
FACES 2000	1609	270	Y	Y	Y
FACES 2003	1522	302	Y	Y	Y
FACES 2006	2075	376	Y	Y	Y
Fragile Families	332	-	N	N	N
Head Start Impact Study	2451	1437	Y	Y	Y
Multistate Study (NCEDL)	830	233	Y	Y	Y
PCER	2340	310	Y	N	Y
QUINCE	179	51	Y	N	Y

Results: Step 1 (Bivariate Models)

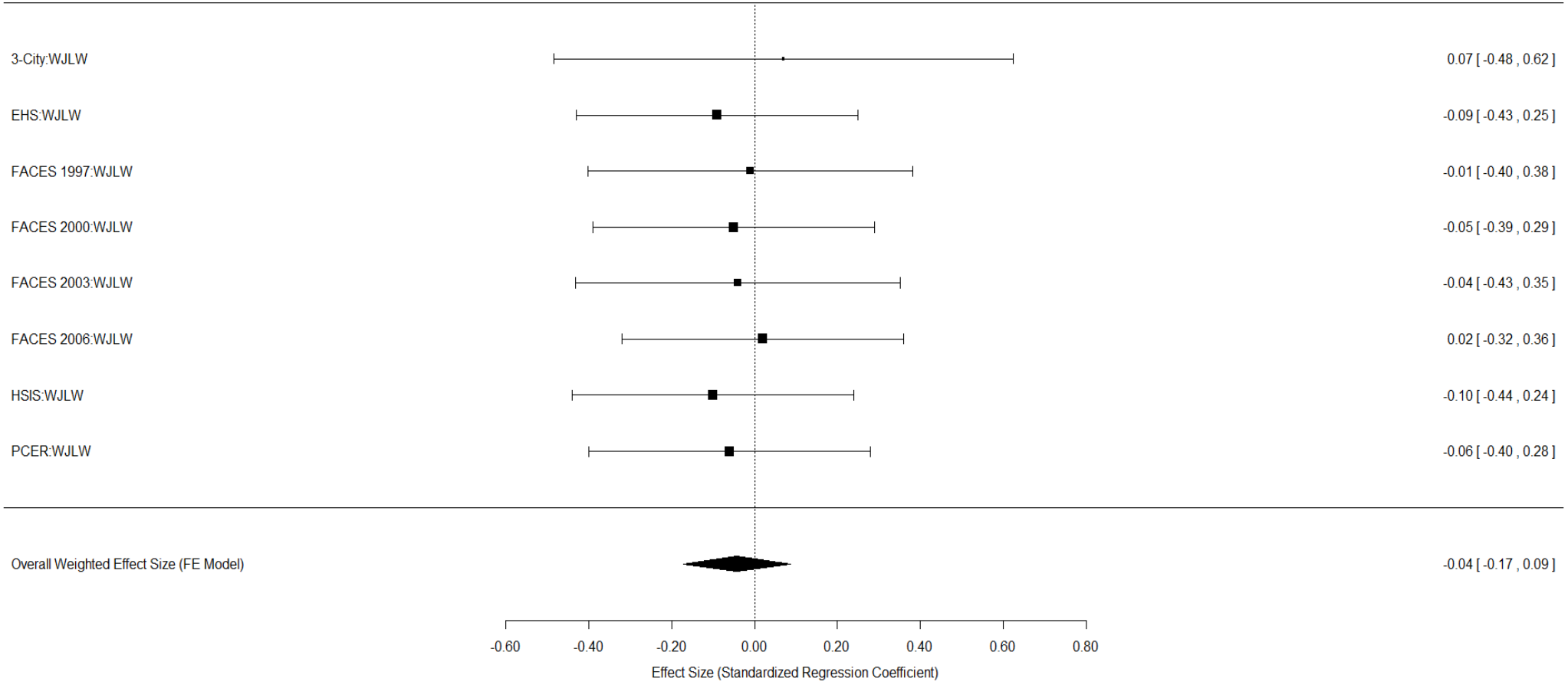
- Across Outcomes:
  - Overall Weighted Effect Size
    - .01 [-.06, .07]
  - Test of Heterogeneity
    - Q (df=35): 3.84, *p*>.05

Predicting Achievement Outcomes from ECERS-R Scores: All Datasets, All Outcomes (Bivariate Models)

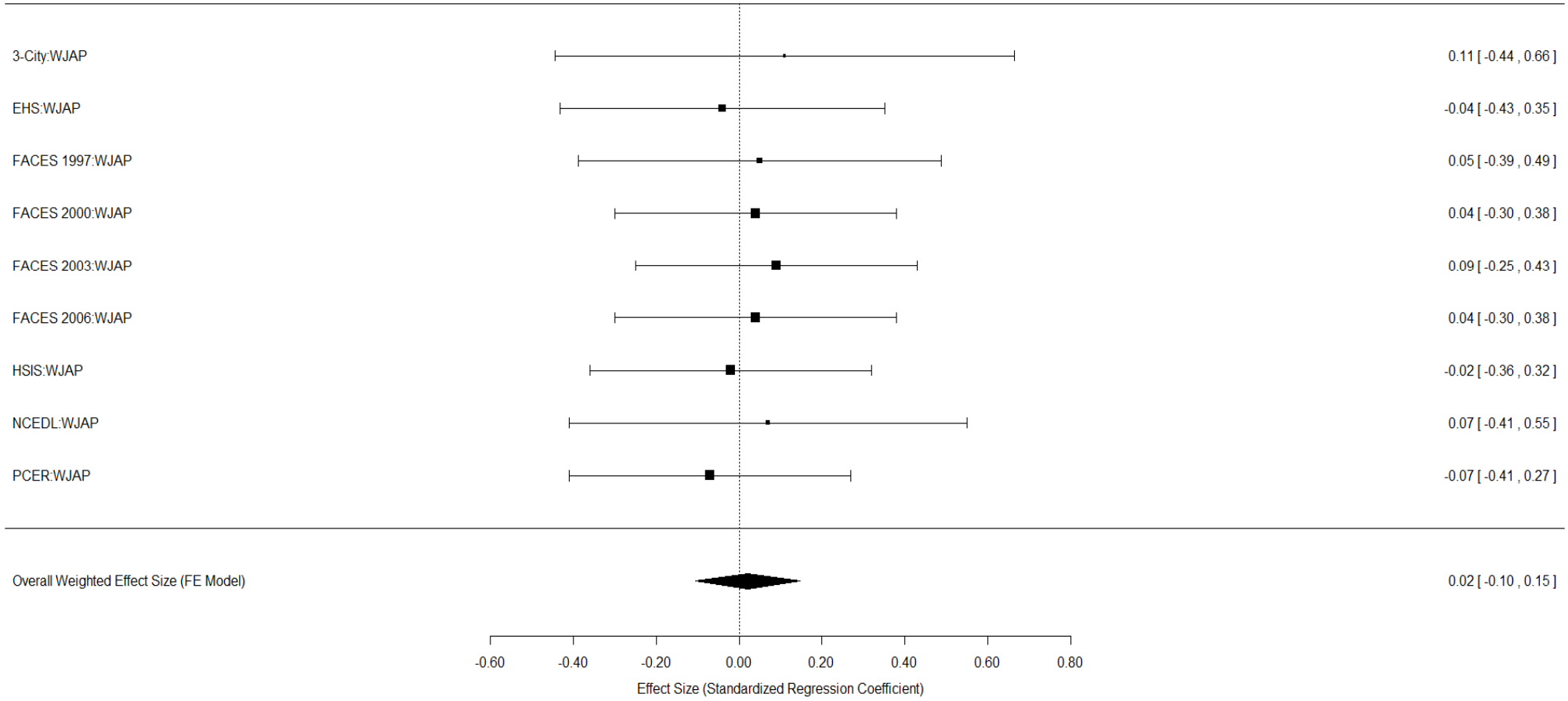



Predicting PPVT from ECERS-R Scores (Bivariate Models)
EHS:PPVT
-0.09 [-0.43 , 0.25]
FACES 1997:PPVT
0.09 [-0.35 , 0.53]
FACES 2000:PPVT
0.12 [-0.32 , 0.56]
FACES 2003:PPVT
0.10 [-0.29 , 0.49]
FACES 2006:PPVT
0.08 [-0.31 , 0.47]
FF:PPVT
0.10 [-0.34 , 0.54]
HSIS:PPVT
0.01 [-0.33 , 0.35]
NCEDL:PPVT
0.16 [-0.32 , 0.64]
PCER:PPVT
0.01 [-0.33 , 0.35]
Overall Weighted Effect Size (FE Model)
0.05 [-0.08 , 0.18]
-0.60
-0.40
-0.20
0.00
0.20
0.40
0.60
0.80
Effect Size (Standardized Regression Coefficient)

Predicting WJ Letter-Word ID from ECERS-R Scores (Bivariate Models)



Predicting WJ Applied Problems from ECERS-R Scores (Bivariate Models)



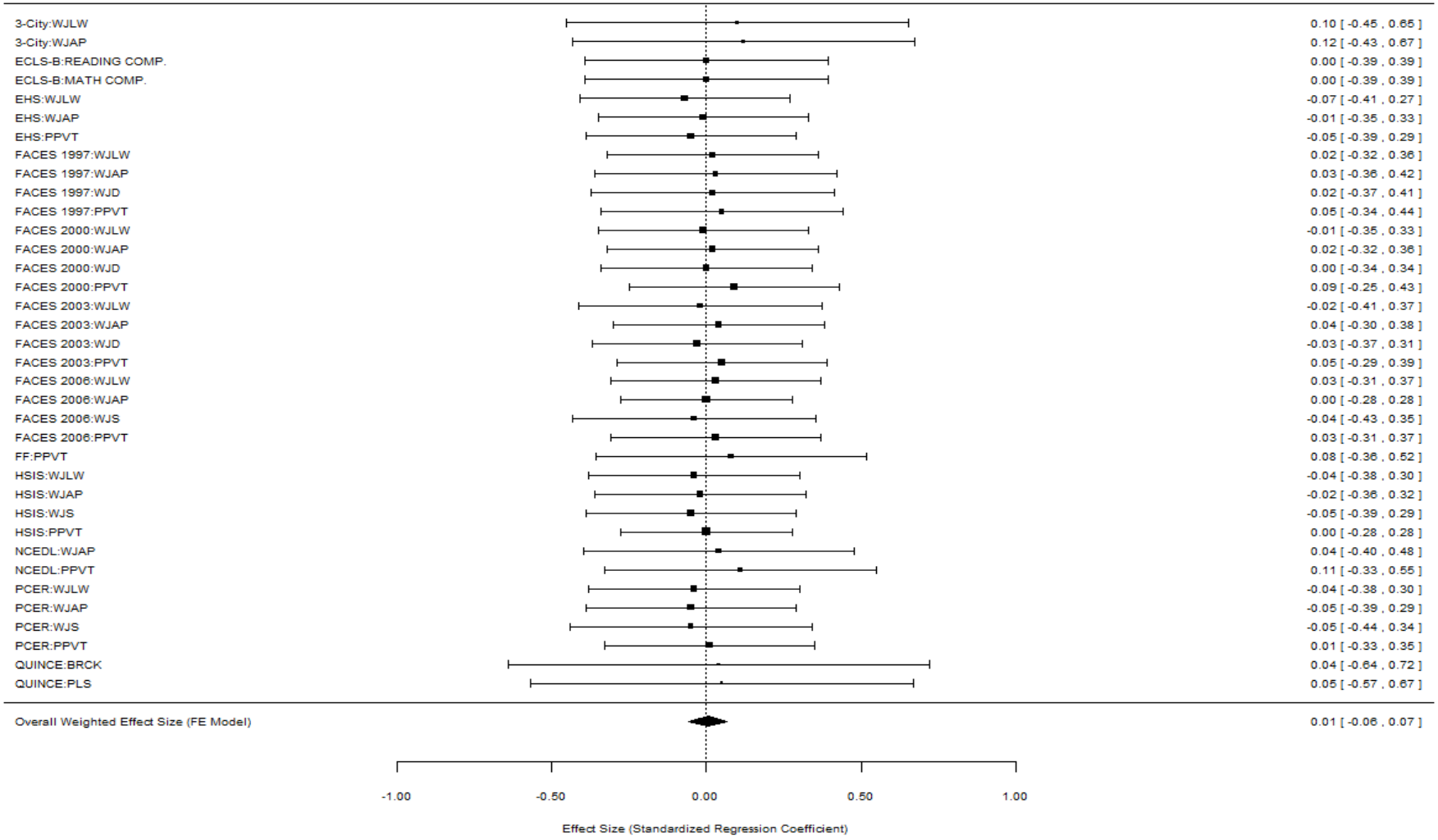
Details: Step 2 (Basic Models)

Study Name	Cohort	Center Type (Head Start)	Center Type (Public School)	Observation Season	Gender	Ethnicity	Age	Low-income Household	Child Disability
3-City Study		X			X	X	X	X	
ECLS-B		X	X	X	X	X	X	X	X
Early Head Start REP					X	X		X	X
FACES 1997				X	X	X	X	X	X
FACES 2000	X			X	X	X	X	X	X
FACES 2003					X	X	X	X	X
FACES 2006					X	X	X	X	X
Fragile Families		X	X	X	X	X	X	X	X
Head Start Impact Study	X	X	X	X	X	X	X	X	X
Multistate Study (NCEDL)		X	X		X	X	X	X	X
PCER									
QUINCE	X	X	X	X	X	X	X	X	

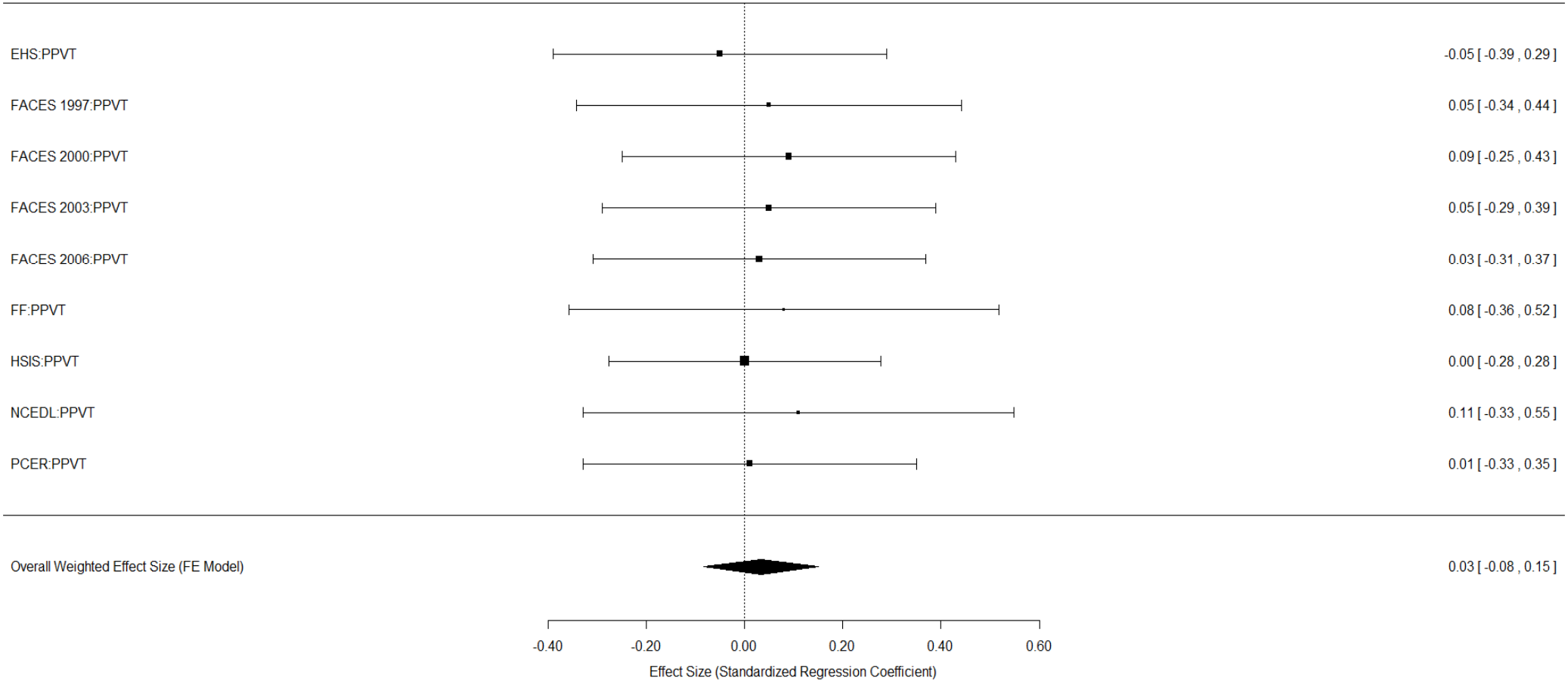
Results: Step 2 (Basic Models)

- Across Outcomes:
  - Overall Weighted Effect Size
    - .01 [-.06, .07]
  - Test of Heterogeneity
    - Q (df=35): 1.99, *p*>.05

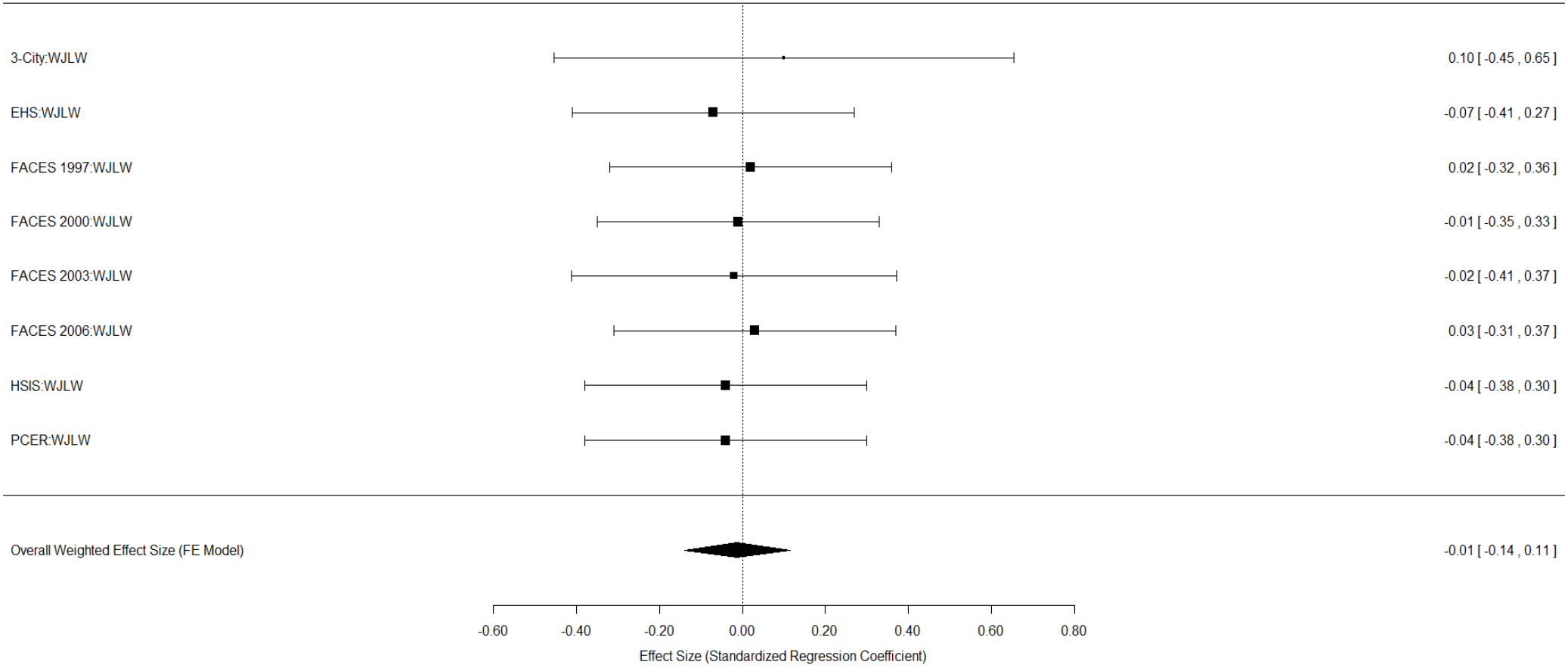
Predicting Achievement Outcomes from ECERS-R Scores: All Datasets, All Outcomes (Basic Models)

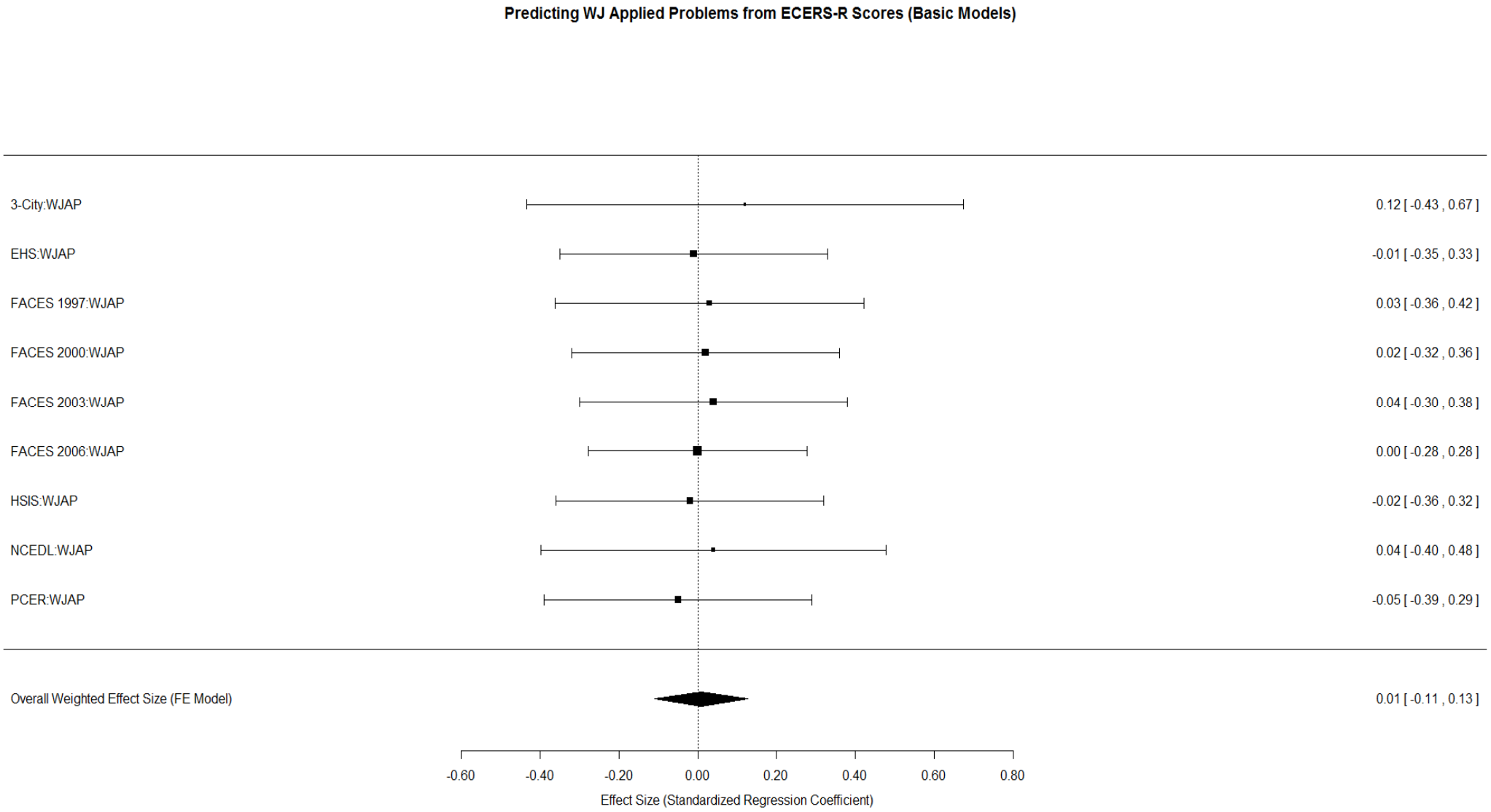


Predicting PPVT from ECERS-R Scores (Basic Models)



Predicting WJ Letter-Word ID from ECERS-R Scores (Basic Models)




Predicting WJ Applied Problems from ECERS-R Scores (Basic Models)
3-City:WJAP
EHS:WJAP
FACES 1997:WJAP
FACES 2000:WJAP
FACES 2003:WJAP
FACES 2006:WJAP
HSIS:WJAP
NCEDL:WJAP
PCER:WJAP
Overall Weighted Effect Size (FE Model)
0.12 [-0.43 , 0.67]
-0.01 [-0.35 , 0.33]
0.03 [-0.36 , 0.42]
0.02 [-0.32 , 0.36]
0.04 [-0.30 , 0.38]
0.00 [-0.28 , 0.28]
-0.02 [-0.36 , 0.32]
0.04 [-0.40 , 0.48]
-0.05 [-0.39 , 0.29]
0.01 [-0.11 , 0.13]
-0.60
-0.40
-0.20
0.00
0.20
0.40
0.60
0.80
Effect Size (Standardized Regression Coefficient)

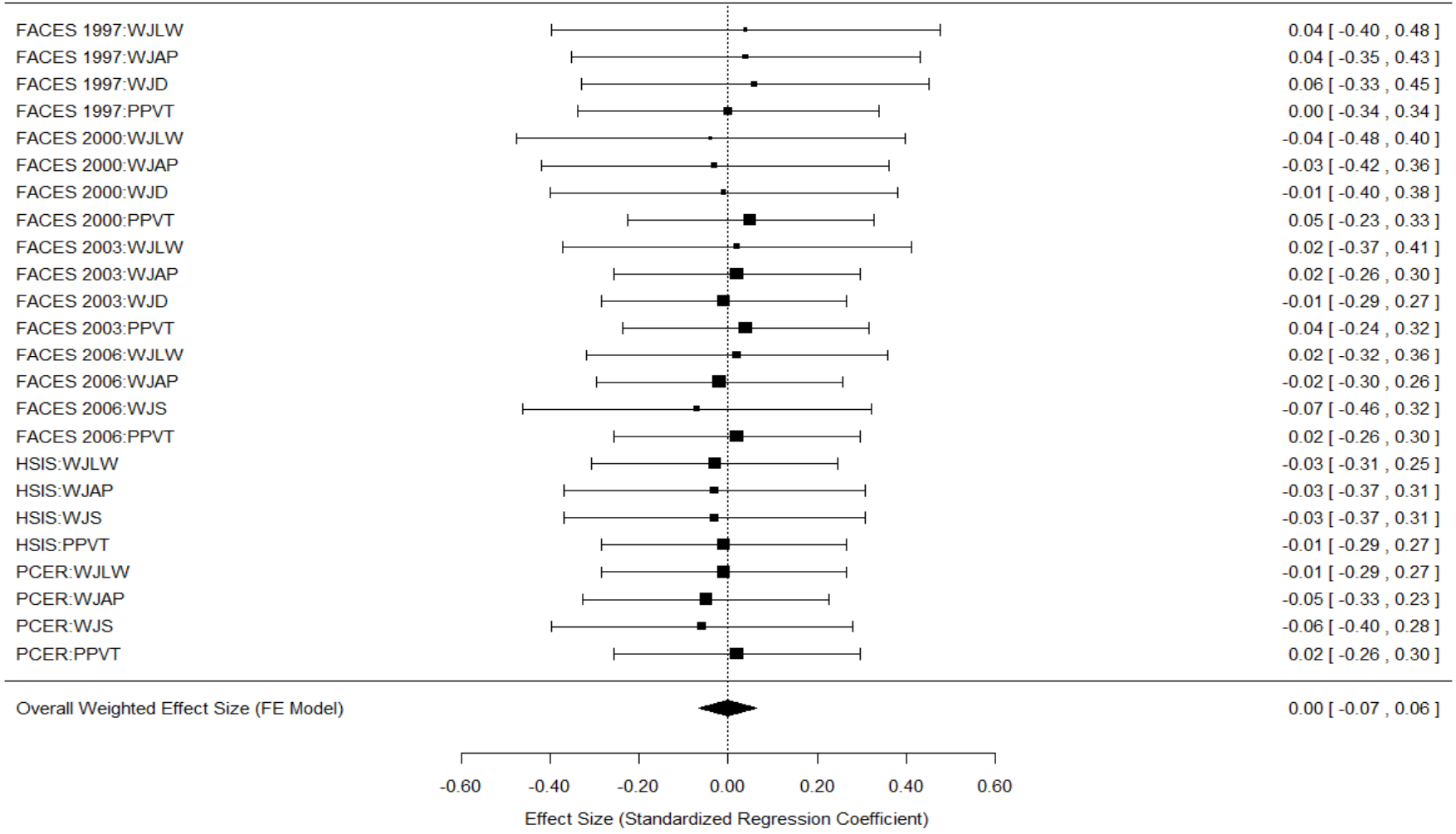
Details: Step 3 (Gain Models)

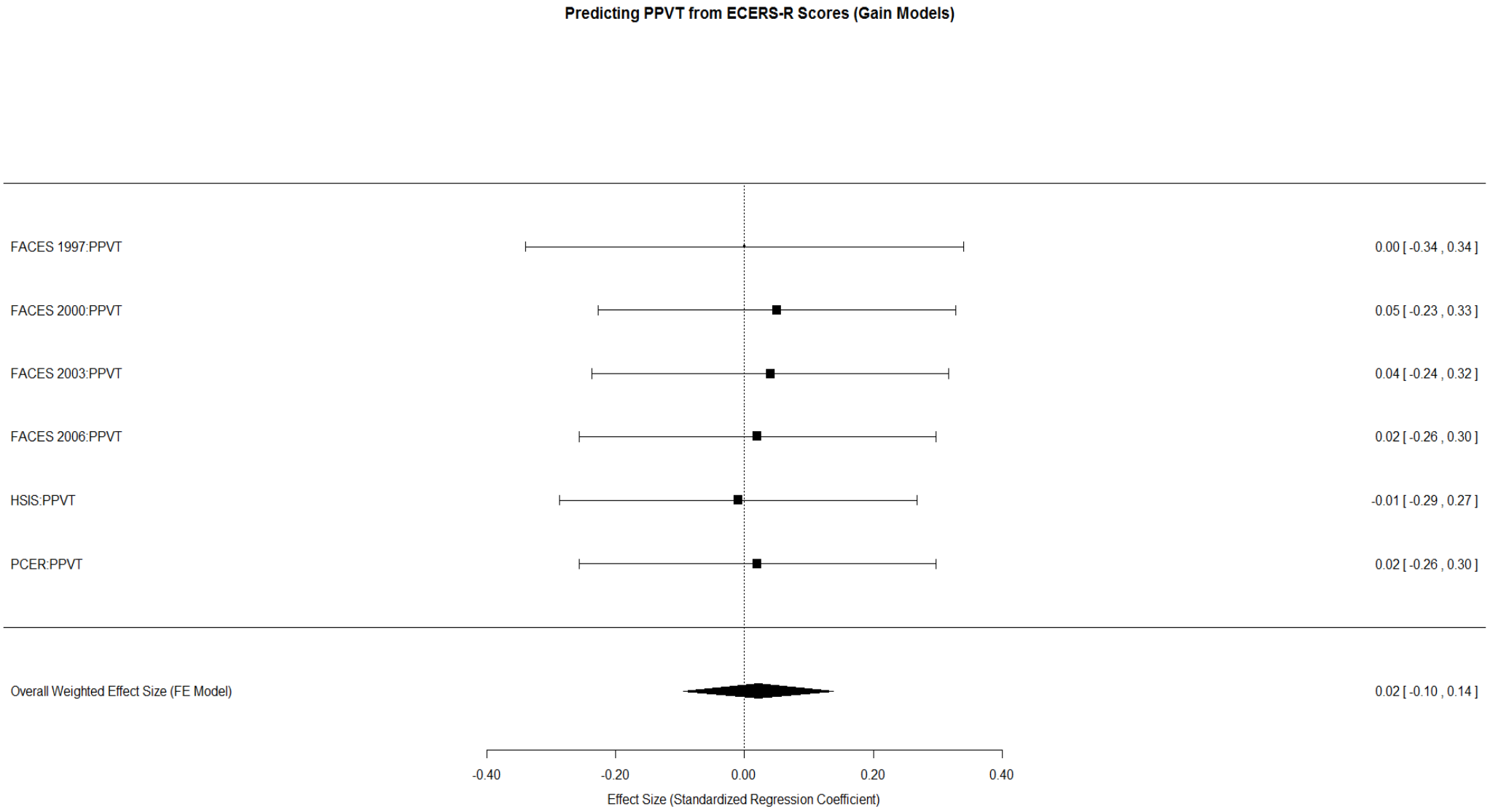
Study Name	Unweighted # of Children	Unweighted # of Classrooms	Multilevel Dataset	Sampling Weight s Used	Robust SE Used
3-City Study					
ECLS-B					
Early Head Start REP					
FACES 1997	531-913	265-357	Y	Y	Y
FACES 2000	593-1506	196-270	Y	Y	Y
FACES 2003	1294	292	Y	Y	Y
FACES 2006	1579	354	Y	Y	Y
Fragile Families					
Head Start Impact Study	2451	1437	Y	Y	Y
Multistate Study (NCEDL)					
PCER	2180	310	Y	N	Y
QUINCE					

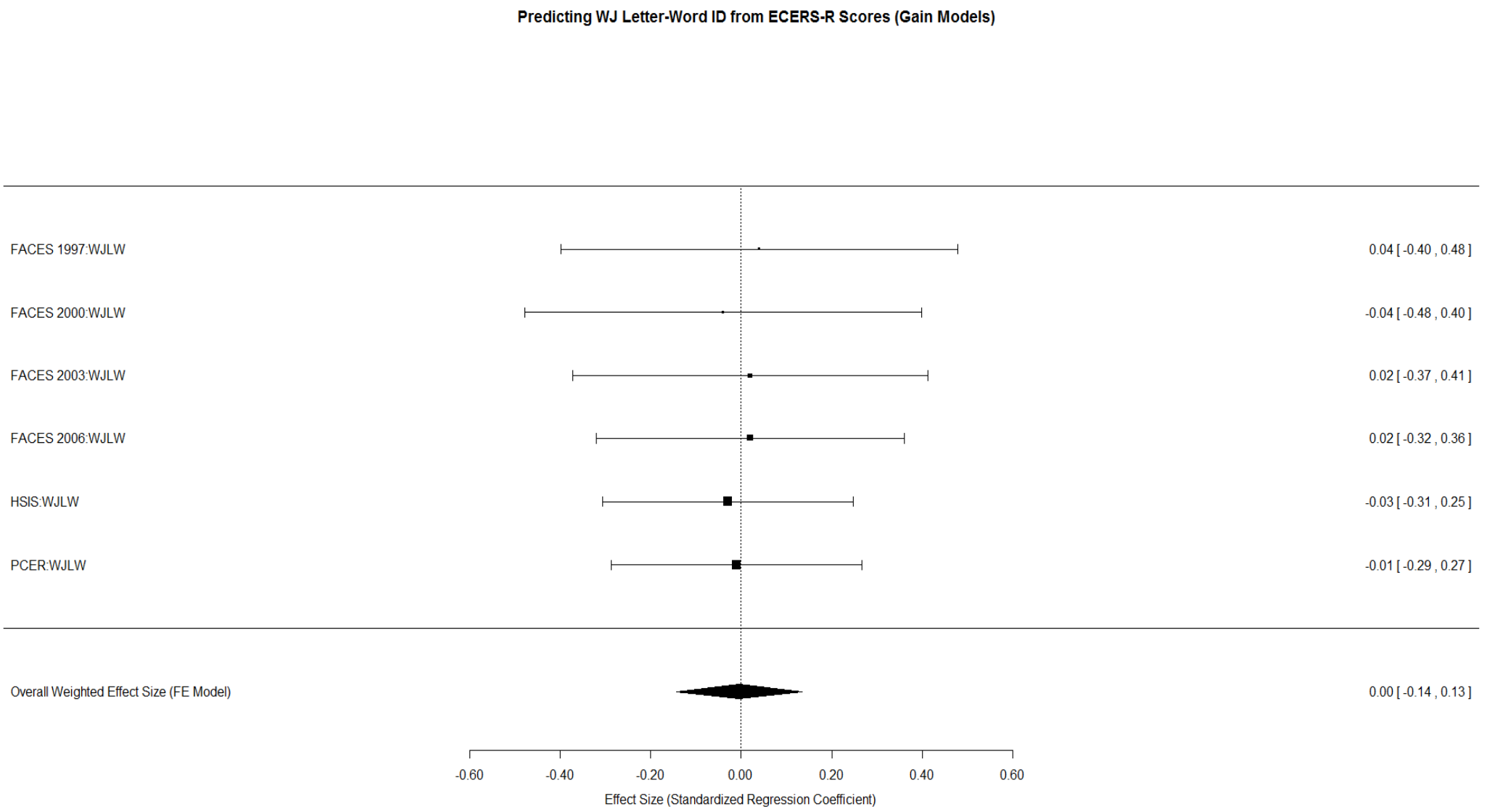
Results: Step 3 (Gain Models)

- Across Outcomes:
  - Overall Weighted Effect Size
    - .00 [-.07, .06]
  - Test of Heterogeneity
    - Q (df=23): 1.01, *p*>.05

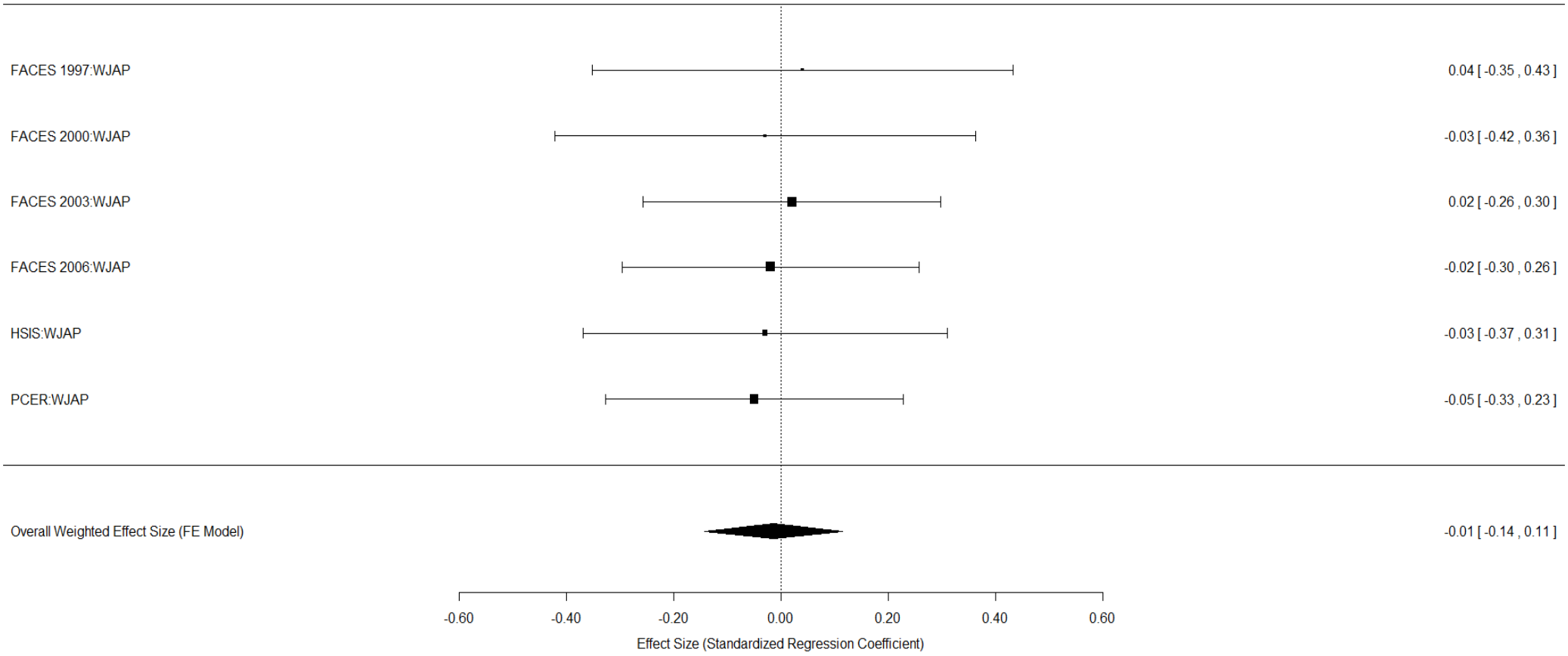
Predicting Achievement Outcomes from ECERS-R Scores: All Datasets, All Outcomes (Gain Models)




Predicting PPVT from ECERS-R Scores (Gain Models)
FACES 1997:PPVT
0.00 [-0.34 , 0.34]
FACES 2000:PPVT
0.05 [-0.23 , 0.33]
FACES 2003:PPVT
0.04 [-0.24 , 0.32]
FACES 2006:PPVT
0.02 [-0.26 , 0.30]
HSIS:PPVT
-0.01 [-0.29 , 0.27]
PCER:PPVT
0.02 [-0.26 , 0.30]
Overall Weighted Effect Size (FE Model)
0.02 [-0.10 , 0.14]
-0.40
-0.20
0.00
0.20
0.40
Effect Size (Standardized Regression Coefficient)


Predicting WJ Letter-Word ID from ECERS-R Scores (Gain Models)
FACES 1997:WJLW
0.04 [-0.40 , 0.48]
FACES 2000:WJLW
-0.04 [-0.48 , 0.40]
FACES 2003:WJLW
0.02 [-0.37 , 0.41]
FACES 2006:WJLW
0.02 [-0.32 , 0.36]
HSIS:WJLW
-0.03 [-0.31 , 0.25]
PCER:WJLW
-0.01 [-0.29 , 0.27]
Overall Weighted Effect Size (FE Model)
0.00 [-0.14 , 0.13]
-0.60
-0.40
-0.20
0.00
0.20
0.40
0.60
Effect Size (Standardized Regression Coefficient)

Predicting WJ Applied Problems from ECERS-R Scores (Gain Models)



Summary: Steps 1-3 (Meta-Analysis)

Model	Weighted E.S.	Q	τ^2
Bivariate			
All Outcomes	0.01	3.84 (*p*>.*10*)	0.00
PPVT	0.05	1.23 (*p*>.*10*)	0.00
WJLW	-0.04	0.51 (*p*>.*10*)	0.00
WJAP	0.02	0.76 (*p*>.*10*)	0.00
Basic			
All Outcomes	0.01	1.99 (*p*>.*10*)	0.00
PPVT	0.03	0.59 (*p*>.*10*)	0.00
WJLW	-0.01	0.42 (*p*>.*10*)	0.00
WJAP	0.01	0.38 (*p*>.*10*)	0.00
Gain			
All Outcomes	0.00	1.01 (*p*>.*10*)	0.00
PPVT	0.02	0.12 (*p*>.*10*)	0.00
WJLW	0.00	0.13 (*p*>.*10*)	0.00
WJAP	-0.01	0.21 (*p*>.*10*)	0.00

Summary: Steps 1-3 (Meta-Analysis)

- These analyses, in all steps and with all specifications, suggest that the relationship between total ECERS-R scores (continuous scale) and achievement outcomes (relatively immediate) is virtually zero.
- Biggest effects were seen for PPVT, but still too close to zero to be practically, let alone statistically, significant.
- There is extremely little variation between datasets that can be attributed to anything other than sampling differences.

ANALYSIS: STEP 4 (USE OF INTEGRATIVE DATA ANALYSIS)

Details: Step 4 (Integrative Data Analysis)

- We also stacked together the datasets, beginning with the four Faces cohorts that have the most parallel designs.

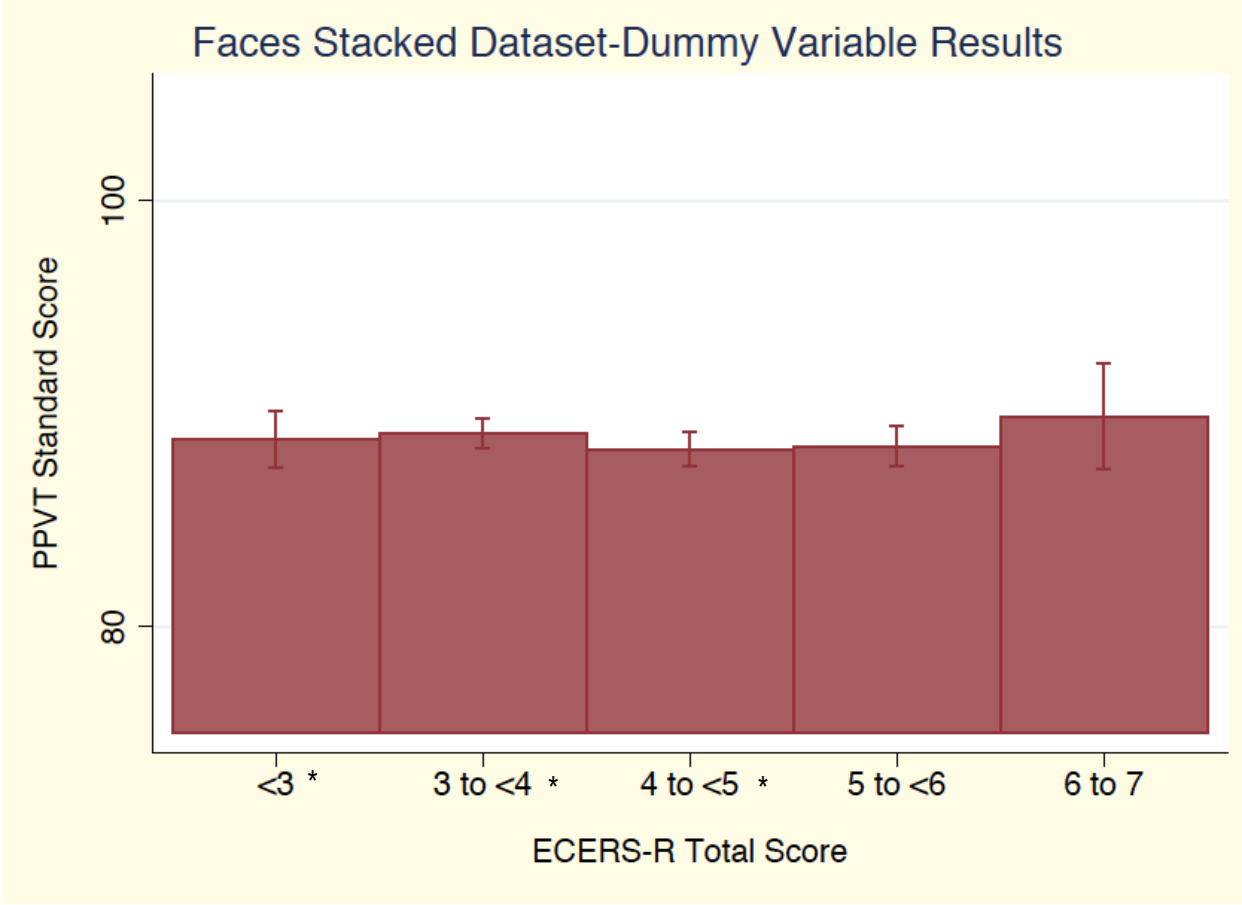
- We first used interactions to confirm that the general (linear) ECERS-R-outcome associations were statistically equivalent across cohorts.

- We then leveraged the larger sample sizes to support tests of non-linear associations (more centers in various regions of the ECERS-R continuum).

Sample Size: Step 4 (Integrative Data Analysis)

ECERS-R Score Grouping	N of Children	N of Classrooms
1 to <3	544	98
3 to <4	2081	396
4 to <5	2036	498
5 to <6	1520	449
6 to <7	319	77
TOTAL	*6,500*	*1,518*

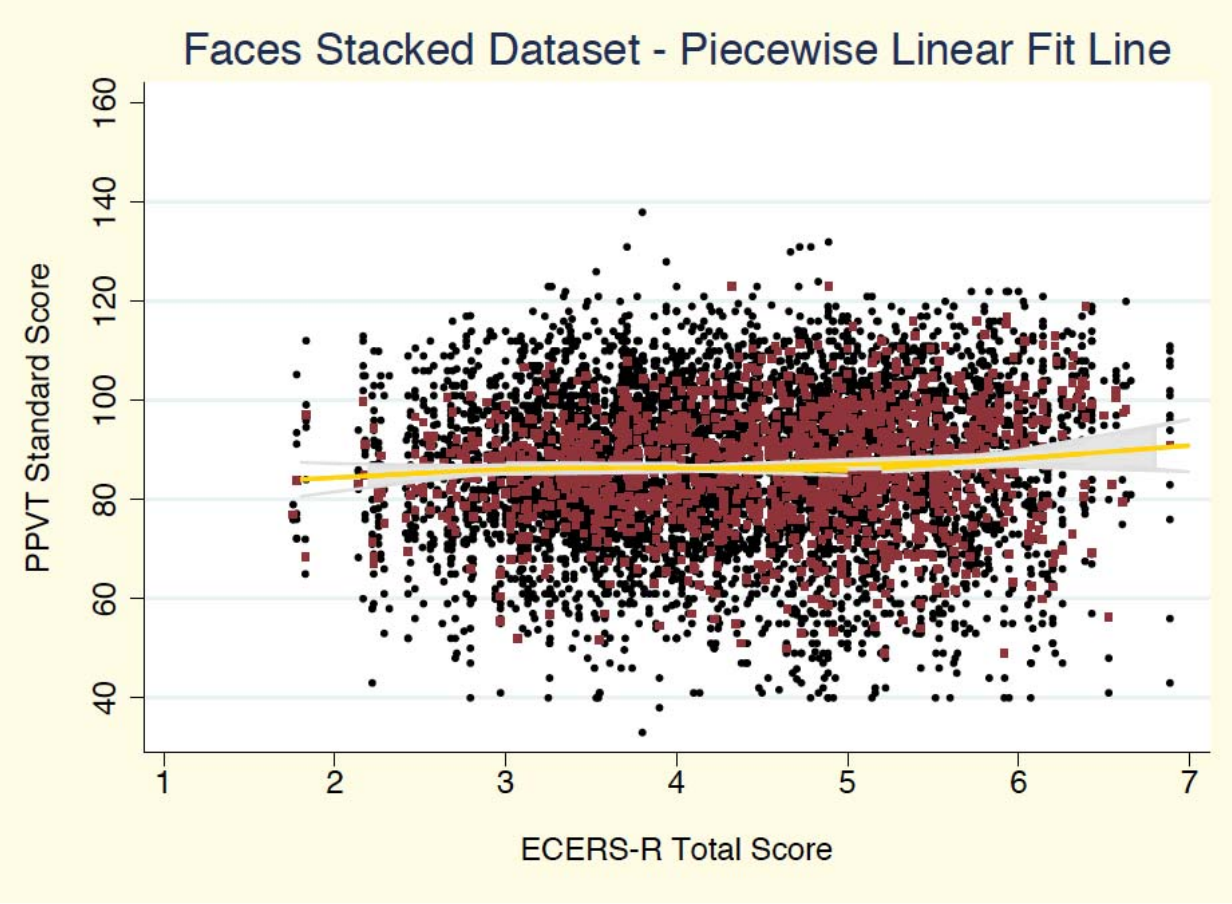
Dummy Variable Results for PPVT: Step 4 (Integrative Data Analysis)



* Significantly different than 6 - 7

- Overall F – Test of any difference in means:
  F (4, 1517) = 1.62, p = .17
- Comparison of specific pairs of means:
  6–7 vs <3 (b = 3.16, *B* = .22
  t = 2.24, p = .03)
  6–7 vs 3-4 (b = 2.47, *B* = .17
  t = 2.05, p = .04)
  6–7 vs 4-5 (b = 2.56, *B* = .18
  t = 2.09, p = .04)

Piecewise Linear Results for PPVT: Step 4 (Integrative Data Analysis)



Slopes within 5 categories:

- The slope is significantly positive between 5 and 6 (b = 2.35, B = .16, p<.05) but not in any other region.

Comparisons of slopes:

- Overall F – Test:
  F (4, 1521) = 1.55
  p = .19
- No slopes within each pair differ significantly.

Summary: Step 4 (Integrative Data Analysis)

- Overall, little evidence of significant non-linearity.
- Somewhat larger associations at the upper end of the quality distribution across the Faces datasets.
- Important to see if these results replicate when we stack the other datasets.

Conclusions

- These preliminary results suggest caution in continued use of the ECERS-R in high stakes policy investments that have the goal of increasing children's scores on the kinds of outcomes examined here.
- In other words, we find little evidence that the ECERS-R helps differentiate between classrooms in which children have higher or lower scores on standardized cognitive assessments.

Next Steps for Analyses

- We plan several next steps for our analyses to further refine our results, which may modify this conclusion including:
  - Examination of different ECERS-R scale transformations
  - Examination of subscale scores (including Language and Reasoning which may align better with cognitive outcomes)
  - Inclusion of different types of outcomes and longitudinal growth
  - Replication of full set of non-linear analyses and MI and FIML missing data analyses across all of the datasets
  - Control for additional covariates, and new methods to synthesize results across datasets and account for missing covariates in some datasets (Wu and Becker, 2013)
  - Additional examination of heterogeneity across subgroups, including subregions/sites in each dataset

Points of Discussion

- The small associations that we find between ECERS-R and outcome scores may reflect additional issues.
  - Does ECERS-R capture aspects of quality that promote children's growth?
  - Would other measures of quality perform better, including those that better focus on cognitive-specific inputs and that differentially capture variation in quality within classrooms (e.g., individual children's experiences)?
  - Do standardized cognitive assessments capture school readiness aspects that are supported by high quality and targeted by policy initiatives?

VANDERBILT
PEABODY COLLEGE

PEABODY research institute

Kerry G. Hofer

E-mail:

kerry.g.hofer@vanderbilt.edu

PRI Website:

http://peabody.vanderbilt.edu/research/pri

Peabody Research Institute

Vanderbilt University

230 Appleton Place, PMB 181

Nashville, TN 37203-5721

(615) 322-0637

(615) 322-0293 (F)