Online Appendix

This supplemental document contains additional information on the math, literacy, social-emotional, and classroom observation measures used across the *Building Blocks* studies. Data alignment considerations that were not addressed in the main text are also included.

Measures

As mentioned in the manuscript, there was variation in the math and non-math assessments used across the five study sites. Measures of math performance at each assessment wave were standardized within study site to allow for the aggregation of data across all five sites. We also generated a composite of non-math measures for each block by standardizing, averaging, and re-standardizing all available measures of non-math skills within a site. Though Table A1 presents a complete list of math and non-math measures implemented across all waves of data collection for each site, we provide a brief description of each measure below. For additional details on each measure, please see the original treatment impact reports (Clements, Sarama, Spitler, et al., 2011; Clements et al., 2020; Hofer et al., 2013; Morris et al., 2016).

Math measures

Woodcock Johnson (III) - Applied Problems. The Woodcock-Johnson Applied Problems (Woodcock et al., 2001) subtest is a direct assessment of simple math functions. Age-normed scores were used, with a mean of 100 and standard deviation of 15.

Research-based Elementary Math Assessment (REMA). The REMA (Clements et al., 2008; Sarama & Clements, 2011) assesses mathematical skills aligned with the *Building Blocks* number (e.g., verbal counting, subitizing, addition and subtraction) and geometry (e.g., shape recognition, construction of shapes, geometric measurement, and patterning) progressions. The measure includes 225 items, which concludes after children make four consecutive errors. Rasch scores are computed for analyses.

Tools for Early Assessment of Mathematics (TEAM). The TEAM (Clements et al., 2008; Clements, Sarama, & Wolfe, 2011) assesses preschool-aged children's mathematical knowledge and skills through two one-on-one interviews with explicit protocols, coding, and scoring procedures.

Early Childhood Longitudinal Study-Birth Cohort (ECLS-B). The ECLS-B (Najarian et al., 2010) is a direct assessment of children's math ability (e.g., geometry, operations, measurement, number sense). Scores range from 0 to 44, with higher scores indicating higher math performance.

Non-Math measures

Renfrew Bus Story – North American Edition. The Renfew Bus Story (Glasgow & Cowley, 1994) is a measure of oral language skills in which children are told a story, asked to recount it based on picture prompts, and answer inferential questions about the story. To align scores across sites, we used the complexity score, utterance length score, and information score.

Woodcock Johnson (III)- Letter Word (WJ-III). The WJ-III Letter Word (Woodcock et al., 2001) subscale assess children's emergent literacy skills using letter recognition, matching, and naming.

Phonological Awareness Literacy Screening (PALS- PreK). The PALS-PreK (Invernizzi et al., 2004) assesses children's early literacy skills, including phonological awareness, word concepts, letter recognition, and letter sounds.

mCLASS:CIRCLE. The mCLASS:CIRCLE (Landry, 2007) is a measure of children's letter recognition and pre-reading skills.

Expressive Vocabulary Test -- Second Edition (EVT-2). The EVT-2 (Williams, 2007) is a test of expressive vocabulary skills in which children have to verbally identify what is represented in a picture.

Receptive One-Word Picture Vocabulary Test, 4th Edition (ROWPVT-4). The ROWPVT-4 (Martin & Brownell, 2011) is a measure of receptive vocabulary skills in which the experimenter says a word and the child has to identify the picture associated with the word.

Peg Tapping. The Peg-Tapping task (Diamond & Taylor, 1996) measures inhibitory control in which children have to inhibit their desire to tap a peg the same amount of times as an assessor.

Head-Toes-Knees-Shoulders (HTKS). The HTKS task (McClelland et al., 2014) measures inhibitory control, attention, and working memory in which children must listen to the experimenter and do the opposite action.

Forward & Backward Digit Span. The Forward & Backward Digit Span task (Carlson, 2016) is a measure of phonological processing and working memory in which children must remember a sequence of numbers and report it backwards.

Self-Ordered Pointing/ Pick the Picture. The Self-Ordered Pointing task (Willoughby et al., 2016) is a measure of working memory. In this task, children view a series of pictures in different arrangements on different pages and select a new picture on each page until all are selected, but none are selected twice.

Item Selection /Something's the Same Game. The Item Selection task (Willoughby et al., 2016) is a measure of attention shifting in which children must identify a match of pictures that share common attributes (e.g., size, color, etc.), with attributes changing across rounds.

Spatial Conflict Arrows. The Spatial Conflict Arrows task (Willoughby et al., 2012) is a measure of cognitive flexibility and inhibition in which children must collect the left button when the arrow points left and the right arrow when the arrow points right, with arrows moving to different parts of the screen (i.e., left arrow moves to the right side).

Corsi Blocks. The Corsi Blocks task (Corsi, 1972; Lezak, 1973) is a measure of short-term memory in which children must recall the arrangement of a series of randomly ordered blocks.

Preschool Self-Regulation Assessment (PSRA). The PSRA (Smith-Donald et al., 2007) is an experimenter-rated assessment of children's attention and impulse control.

Classroom Quality measures

Classroom quality was measured using the Classroom Observation of Early Mathematics—Environment and Teaching (COEMET; e.g., Clarke & Clarke, 2004; Clements et al., 2004), a 3-hour observational assessment of children's math environment. In addition to the information mentioned in the manuscript, there were a few notable alignment issues that should be noted. First, in NYC, an abbreviated version of the COEMET was administered that involved 7 items of math instruction, such as whether each math activity included: teacher involvement, encouragement sharing, encouraged listening, supported understanding, provided support, built on child ideas, and encouraged reflection. In order to generate comparable quality measures across the five sites, we only considered these 7 items in the other 4 sites. Prior to generating our average measure of math instructional quality, we observed a negative correlation between teacher involvement and the other 6 items. The NYC COEMET also relied on different Likert scale anchors than the other four sites. Thus, we dichotomized the continuous quality ratings to be consistent across the 5 sites (1=agree/strongly agree, 0=all else). We then used these 6 binary (i.e., dropping teacher involvement) ratings to create the average math quality measure (i.e., COEMET).

There were also a few issues related to the alignment of the time spent on math activities. To create the proportion of time spent on math activities, the observation length was needed.

This information was not available for the Tennessee site. Therefore, the total observational period was set to 3 hours in Tennessee, which is the intended observation duration of the COEMET.

Missingness

To address missingness we used mean imputation within each block for the following variables: gender, race/ethnicity, language proficiency, age at pretest, age at posttest, and pretest math assessment. While missingness was generally minimal, there was considerable missingness observed in San Diego and NYC. In San Diego, 165 of 699 participants had information on race/ethnicity, limited English proficiency status, and 534 participants had data on age at posttest. In NYC, 714 of 1216 participants had data on pre-test age and pre-test math ability. A "dummy" variable was created to indicate whether imputation was used for each variable and was controlled for in the affected models.

Supplemental References

Carlson, S. M. (2016). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595–616.

Clarke, D. M., & Clarke, B. (2004). Mathematics teaching in grades K-2: Painting a picture of challenging, supportive, and effective classrooms. In R. N. Rubenstein & G. W. Bright (Eds.), *Perspectives on the teaching of mathematics. Sixty-sixth yearbook* (pp. 67–81). National Council of Teachers of Mathematics.

Clements, D. H., Sarama, J., Layzer, C., Unlu, F., & Fesler, L. (2020). Effects on mathematics and executive function of a mathematics and play intervention versus mathematics alone. *Journal for Research in Mathematics Education*, *51*(3), 301–333.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education, 42*(2), 127–166.

Clements, D. H., Sarama, J., & Wolfe, C. B. (2011). *TEAM—Tools for early assessment in mathematics*. McGraw-Hill Education.

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-Based Early Maths Assessment. *Educational Psychology, 28*(4), 457–482.

Clements, D. H., & Stephan, M. (2004). Measurement in pre-K to grade 2 mathematics. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 299–317). Routledge.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* [Doctoral dissertation, McGill University].

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do." *Developmental Psychobiology, 29*(4), 315–334.

Glasgow, C., & Cowley, J. (1994). *Renfrew Bus Story test* (North American Edition). Centreville School.

Hofer, K. G., Lipsey, M. W., Dong, N., & Farran, D. C. (2013). *Results of the Early Math Project – Scale-up cross-site results. Working Paper.* Peabody Research Institute.

Invernizzi, M., Sullivan, A., Swank, L., & Meier, J. (2004). *Phonological awareness literacy screening for preschoolers (PALS pre-K)* (2nd ed.). University of Virginia.

Landry, S. (2007). *mClass:CIRCLE.* Wireless Generation.

Lezak, M. D. (1983). *Neuropsychological Assessment*. Oxford University Press.

Martin, N. A., & Brownell, R. (2011). *Receptive One-Word Picture Vocabulary Test — Fourth Edition*. Academic Therapy Publications.

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The Head-Toes-Knees-Shoulders task. *Frontiers in Psychology, 5*, 599.

Morris, P. A., Mattera, S. K., & Maier, M. F. (2016). *Making pre-K count: Improving math instruction in New York City*. MDRC.

Najarian, M., Snow, K., Lennon, J., Kinsey, S., & Mulligan, G. (2010). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) Preschool—Kindergarten 2007 Psychometric*

Report (No. NCES 2010-009). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, *22*(2), 173–187.

Williams, K. T. (2007). *EVT-2: Expressive vocabulary test, second edition.* Form B. Pearson Education.

Willoughby, M. T., Blair, C. B., & The Family Life Project Investigators. (2016). Measuring executive function in early childhood: A case for formative measurement. *Psychological Assessment, 28*(3), 319–330. https://doi.org/10.1037/pas0000152

Willoughby, M. T., Wirth, R. J., Blair, C. B., & Family Life Project Investigators. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment 24*(2), 418–431.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Riverside Publishing.

Table A1

Measures Used Across Study Sites

	Buffalo	Boston	Tennessee	San Diego	NYC
Math Measures					
Research-based Elementary Math Assessment (REMA)	PK Fall, PK Spring, K Spring	PK Fall, PK Spring, K Spring	PK Fall, PK Spring, K Spring		K Spring
Woodcock Johnson (III) - Applied Problems (WJ-AP)			PK Fall, PK Spring, K Spring		PK Spring, K Spring
Tools for Early Assessment of Mathematics (TEAM)				PK Fall, PK Spring, K Spring	
Early Childhood Longitudinal Study-Birth Cohort (ECLS-B)				PK Spring, K Spring	PK Fall, PK Spring
Classroom Observation of Early Mathematics Environment and Teaching (COEMET)	PK Spring	PK Spring	PK Spring	PK Spring	PK Spring
Non-Math Measures					
Language and Literacy					
Renfrew Bus Story	K Fall	K Fall	PK Spring	PK Spring	
Woodcock Johnson (III) - Letter Word			PK Spring		
Phonological Awareness Literacy Screening (PALS) [a.]		PK Spring		PK Spring	
mCLASS:CIRCLE	PK Spring				
Expressive Vocabulary Test, second edition (EVT-2)				PK Spring	
Receptive One Word Picture Vocabulary Test, 4th edition (ROWPVT-4)					PK Spring

Executive Function & Behavioral Measures		
Peg Tapping	PK Spring	
Head-Toes-Knees-Shoulders (HTKS)	PK Spring	
Forward & Backward Digit Span	PK Spring	
Self-Ordered Pointing	PK Spring	
Item Selection	PK Spring	
Spatial Conflict Arrows		PK Spring
Corsi Blocks (Forwards & Backwards)		PK Spring
Preschool Self-Regulation Assessment (PSRA) attention/impulsivity control		PK Spring

Note. "PK" = Pre-Kindergarten; "K" = Kindergarten.

a. Boston used the Alphabet Knowledge Subtest; San Diego used both Alphabet Knowledge and Name Writing Subtests.

Table A2

Pairwise Correlation Between Within-Site Math Measures

Buffalo					
	REMA Pre-K Entry	REMA Pre-K Post			
REMA Pre-K Entry	1.00				
REMA Pre-K Post	0.60***	1.00			
REMA K	0.59***	0.77***			
Observations	521				
Boston					
	REMA Pre-K Entry	REMA Pre-K Post			
REMA Pre-K Entry	1.00				
REMA Pre-K Post	0.58***	1.00			
REMA K	0.53***	0.73***			
Observations	174				
Tennessee					
	REMA Pre-K Entry	WJ-AP Pre-K Entry	REMA Pre-K Post	WJ-AP Pre-K Post	REMA K
REMA Pre-K Entry	1.00				
WJ-AP Pre-K Entry	0.57***	1.00			
REMA Pre-K Post	0.62***	0.61***	1.00		
WJ AP Pre-K Post	0.55***	0.63***	0.70***	1.00	
REMA K	0.58***	0.57***	0.73***	0.67***	1.00
WJ-AP K	0.55***	0.55***	0.63***	0.61***	0.75***
Observations	771				
San Diego					
	TEAM Pre-K Entry	TEAM Pre-K Post	ECLSB Pre-K Post	TEAM K	
TEAM Pre-K Entry	1.00				
TEAM Pre-K Post	0.68***	1.00			
ECLSB Pre-K Post	0.67***	0.71***	1.00		
TEAM K	0.67***	0.72***	0.75***	1.00	
ECLSB K	0.56***	0.63***	0.67***	0.73***	
Observations	695				
NYC					
	ECLSB Pre-K Entry	ECLSB Pre-K Post	WJ AP-Pre-K Post	WJ-AP K	
ECLSB Pre-K Entry	1.00				
ECLSB Pre-K Post	0.73***	1.00			
WJ AP-Pre-K Post	0.60***	0.71***	1.00		
WJ-AP K	0.54***	0.63***	0.65***	1.00	
REMA K	0.53***	0.63***	0.61***	0.69***	
Observations	1216				

Note. + $p < 0.1$ *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Table A3

Percent Missing Data on Baseline Characteristics Across Sites

	Buffalo	Boston	Tennessee	San Diego	NYC
Female	0%	0%	0%	1.43%	0%
Black- Non Hispanic	0%	0%	0%	0%	0%
White- Non Hispanic	0%	0%	0.65%	76.39%	2.14%
Ethnicity- Other	0%	0%	0.65%	76.39%	2.14%
Hispanic	0%	0%	0.65%	76.39%	2.14%
Limited Eng Prof.	0%	0%	0.26%	76.39%	0.33%
Age at Pre-K Entry (years)	0%	0%	0.40%	4.58%	41.28%
Age at Pre-K Post (years)	0.19%	0%	10.64%	23.61%	0%
Pre-Test Math (Std)	0%	0%	0.52%	3.29%	41.28%

Table A4

Child-Level Descriptive Statistics by Experimental Condition Across Study Sites

	Buffalo			**Boston**			**Tennessee**			**San Diego**			**NYC**		
	Tx *M (SD)*	**Control** *M (SD)*	*p-value*	**Tx** *M (SD)*	**Control** *M (SD)*	*p-value*	**Tx** *M (SD)*	**Control** *M (SD)*	*p-value*	**Tx** *M (SD)*	**Control** *M (SD)*	*p-value*	**Tx** *M (SD)*	**Control** *M (SD)*	*p-value*
Female	0.51	0.52	0.717	0.48	0.49	0.843	0.54	0.55	0.706	0.57	0.51	0.0597+	0.52	0.53	0.497
Black – Non Hispanic	0.55	0.58	0.980	0.44	0.35	0.629	0.8	0.73	0.451	0.06	0.04	0.480	0.34	0.35	0.382
White – Non Hispanic	0.33	0.17	0.243	0.08	0.14	0.923	0.08	0.12	0.285	0.25	0.24	0.952	0.06	0.01	0.053+
Ethnicity- Other	0.03	0.03	0.757	0.08	0.1	0.105	0.04	0.05	0.524	0.18	0.26	0.298	0.03	0.04	0.185
Hispanic	0.09	0.22	0.223	0.39	0.41	0.345	0.08	0.10	0.979	0.51	0.46	0.390	0.57	0.59	0.715
Limited Eng Prof.	0.02	0.12	0.177	0.32	0.39	0.216	0.09	0.11	0.913	0.4	0.47	0.419	0.08	0.12	0.106
Age at Pre-K Entry (years)	4.21 (0.29)	4.31 (0.32)	0.000***	4.6 (0.31)	4.68 (0.29)	0.006**	4.43 (0.31)	4.48 (0.31)	0.013*	4.46 (0.35)	4.33 (0.43)	0.009**	4.29 (0.29)	4.31 (0.29)	0.505
Age at Pre-K Post (years)	4.92 (0.28)	4.92 (0.32)	0.971	5.3 (0.32)	5.32 (0.32)	0.137	5.04 (0.32)	5.07 (0.30)	0.125	5.00 (0.43)	4.98 (0.33)	0.505	4.81 (0.30)	4.84 (0.30)	0.092+
Pre-Test Math Assessment (z-score)	-0.11 (1.16)	0.00 (1.00)	0.435	-0.09 (0.91)	0.00 (1.00)	0.462	0.16 (0.96)	0.00 (1.00)	0.013*	0.28 (0.91)	0.00 (1.00)	0.030*	0.32 (0.98)	0.00 (1.00)	0.000***
Observations	296	225	521	95	79	174	452	319	771	333	366	699	525	691	1216

Note. *$p < .05$. **$p < .01$. ***$p < .001$. Standard deviations are provided in parentheses. *P*-values were generated by regressing each baseline characteristic on the treatment indicator for site, with standard errors adjusted for school site clustering and blocking group included as a control. "Tx" = Treatment

Table A5

Correlations Between Key Predictor Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. Post-Test Impacts (Adj for Pre-Test)	1.00												
2. Follow-Up Impacts (Adj for Pre-Test)	0.56***	1.00											
3. Total Post-Test Level	0.22	0.08	1.00										
4. Treatment Group Post-Test Level	0.53***	0.33*	0.87***	1.00									
5. Control Group Post-Test Level	-0.19	-0.22	0.84***	0.46**	1.00								
6. Total Gains Between Pre-K and K	-0.33*	-0.32*	-0.61***	-0.56***	-0.47**	1.00							
7. Treatment Group Gains Between Pre-K and K	-0.58***	-0.31*	-0.57***	-0.71***	-0.24	0.83***	1.00						
8. Control Group Gains Between Pre-K and K	0.05	-0.22	-0.43**	-0.21	-0.55***	0.82***	0.37*	1.00					
9. Non-Math Cognitive Impacts (Pre-K; Adj. for Pre-Test)	0.29+	-0.08	-0.07	-0.01	-0.12	-0.02	-0.08	0.04	1.00				
10. Impact on Overall COEMET Quality Score (Pre-K)	0.28+	0.21	-0.16	0.04	-0.33*	-0.25	-0.34*	-0.04	-0.09	1.00			
11. Impact on Math Instructional Time (Pre-K)	0.20	-0.18	0.16	0.18	0.08	0.04	-0.16	0.25	-0.12	0.09	1.00		
12. Impact on Number of Math Activities (Pre-K)	0.34*	0.19	0.25	0.42**	-0.01	-0.29+	-0.41**	-0.04	0.02	0.31*	0.18	1.00	
13. Black Non-Hispanic	0.29+	0.02	0.11	0.20	-0.02	-0.00	-0.21	0.21	0.09	0.18	0.14	0.35*	1.00
14. Limited English Proficiency	-0.08	0.11	-0.28+	-0.20	-0.28+	-0.21	-0.16	-0.18	-0.25	0.21	0.10	-0.14	-0.57***
Observations	41												

Note. + $p < 0.1$ *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Table A6

Econometric Fixed Meta-Regressions

	FE Model-Baseline	FE Model-Gains
	(1)	(2)
Post-Test Impact (Adj. For Pre-Test)	0.42**	0.28
	(0.15)	(0.23)
Control Group Post-Test Level		-0.64*
		(0.28)
Control Group Gains Between Prek-Post and K		-0.97**
		(0.37)
Treatment Group Post-Test Level		0.48
		(0.30)
Treatment Group Gains Between Prek-Post and K		0.56
		(0.40)
N (Site / Study Blocks)	5 / 41	5 / 41

Note. *$p < .05$. **$p < .01$. ***$p < .001$. Standard errors are provided in parentheses. All models are run at the block level (n=41) and control for children's pre-test assessment. Models rely on a fixed effects meta-analytic model with an econometric fixed effect for city and weights.

Table A7

Econometric Fixed Meta-Regressions Exploring Predictors of Follow-Up Impact

	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts (Adj for Pre-Test)	0.55**	0.45**	0.46**	0.45**	0.44**	0.41**
	(0.17)	(0.17)	(0.16)	(0.17)	(0.16)	(0.16)
Non-Math Cognitive Impacts (Pre-K; Adj. for Pre-Test)	-0.28					
	(0.17)					
Impact on Overall COEMET Quality Score (Pre-K)		0.00				
		(0.05)				
Impact on Math Instructional Time (Pre-K)			-0.05			
			(0.04)			
Impact on Number of Math Activities (Pre-K)				0.00		
				(0.05)		
Limited English Proficiency (10%)					0.02	
					(0.03)	
Black-Non Hispanic (10%)						0.00
						(0.02)
N (Site / Study Blocks)	5 / 41	5 / 40	5 / 40	5 / 40	5 / 41	5 / 41

Note. *$p < .05$. **$p < .01$. ***$p < .001$. Standard errors are provided in parentheses. All models are run at the block level (n=41) and control for children's pre-test assessment. All models rely on a fixed effects meta-analytic model with an econometric fixed effect for city and weights. Both Limited English Proficiency and Black Non-Hispanic variables were rescaled to 10 percentage point units.

Table A8
Predicting Follow-up Impacts without Pre-Test Adjustments

	(1)	(2)	(3)	(4)
Post-Test Impacts	0.43**	0.44**	0.46**	0.59***
	(0.13)	(0.14)	(0.16)	(0.16)
Total Post-Test Level		-0.21		
		(0.17)		
Total Gains Between Pre-K Post and K		-0.29		
		(0.31)		
Control Group Post-Test Level			-0.33*	
			(0.17)	
Control Group Gains Between Pre-K Post and K			-0.65*	
			(0.27)	
Treatment Group Post-Test Level				0.00
				(0.18)
Treatment Group Gains Between Pre-K Post and K				0.40
				(0.33)
Constant	0.04	0.40	0.86*	-0.48
	(0.06)	(0.40)	(0.35)	(0.42)
N (City / Study Blocks)	5 / 41	5 / 41	5 / 41	5 / 41
τ (site)	0.01	0.06	0.08	0.11

Table A9

Exploratory Predictors of Follow-Up Impact Without Pre-test Adjustments

	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts	0.43**	0.44***	0.47***	0.44**	0.45***	0.50***
	(0.14)	(0.14)	(0.14)	(0.14)	(0.13)	(0.14)
Non-Math Cognitive Impacts (Pre-K)	0.02					
	(0.18)					
Impact on Overall COEMET Quality Score (Pre-K)		0.00				
		(0.06)				
Impact on Math Instructional Time (Pre-K)			-0.09			
			(0.08)			
Impact on Number of Math Activities (Pre-K)				-0.01		
				(0.07)		
Limited English Proficiency (10%)					0.04+	
					(0.02)	
Black-Non Hispanic (10%)						-0.02
						(0.02)
Constant	0.04	0.04	0.11	0.04	-0.04	0.12
	(0.06)	(0.06)	(0.08)	(0.07)	(0.07)	(0.08)
N (City / Study Blocks)	5 / 41	5 / 40	5 / 40	5 / 40	5 / 41	5 / 41
τ (site)	0.03	0.03	0.00	0.01	0.00	0.00

Note. *$p < .05$. **$p < .01$. ***$p < .001$. Standard deviations are provided in parentheses. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site and weights. All models are run at the block level (n=41) and do not adjust for pre-test performance. Both Limited English Proficiency and Black Non-Hispanic variables were rescaled to 10 percentage point units.

Table A10

Predicting Follow-Up Impact with Each Study Site Dropped Sequentially

	Buffalo Dropped				
	(1)	(2)	(3)	(4)	(5)
Post-Test Impacts (Adj for Pre-Test)	0.33*	0.27+	0.35*	0.25	0.21
	(0.14)	(0.15)	(0.14)	(0.19)	(0.25)
Total Post-Test Level		-0.14			
		(0.15)			
Total Gains Between Pre-K Post and K		-0.56*			
		(0.26)			
Control Group Post-Test Level			-0.21		-0.69*
			(0.14)		(0.30)
Control Group Gains Pre-K Post to K			-0.56**		-0.90**
			(0.20)		(0.29)
Treatment Group Post-Test Level				0.00	0.59+
				(0.16)	(0.33)
Treatment Group Gains Pre-K Post to K				-0.18	0.62
				(0.32)	(0.40)
Constant	0.00	0.73*	0.75**	0.22	0.38
	(0.06)	(0.34)	(0.27)	(0.41)	(0.37)
N (City / Study Blocks)	4 / 35	4 / 35	4 / 35	4 / 35	4 / 35
τ (site)	0.07	0.00	0.00	0.06	0.00
	Boston Dropped				
	(1)	(2)	(3)	(4)	(5)
Post-Test Impacts (Adj for Pre-Test)	0.38**	0.38**	0.43***	0.42**	0.37
	(0.13)	(0.13)	(0.12)	(0.16)	(0.24)
Total Post-Test Level		-0.14			
		(0.14)			
Total Gains Between Pre-K Post and K		-0.38			
		(0.25)			
Control Group Post-Test Level			-0.18		-0.56
			(0.13)		(0.35)
Control Group Gains Pre-K Post to K			-0.46*		-0.87**
			(0.20)		(0.32)
Treatment Group Post-Test Level				-0.02	0.46
				(0.15)	(0.37)
Treatment Group Gains Pre-K Post to K				0.08	0.65+
				(0.31)	(0.37)
Constant	0.02	0.52	0.63*	-0.08	0.31
	(0.06)	(0.32)	(0.27)	(0.40)	(0.34)
N (City / Study Blocks)	4 / 37	4 / 37	4 / 37	4 / 37	4 / 37
τ (site)	0.07	0.00	0.00	0.09	0.00

	Tennessee Dropped				
	(1)	(2)	(3)	(4)	(5)
Post-Test Impacts (Adj for Pre-Test)	0.54***	0.55***	0.56***	0.53**	0.41+
	(0.14)	(0.15)	(0.14)	(0.17)	(0.24)
Total Post-Test Level		-0.20			
		(0.15)			
Total Gains Between Pre-K Post and K		-0.46+			
		(0.26)			
Control Group Post-Test Level			-0.25+		-0.59*
			(0.13)		(0.27)
Control Group Gains Pre-K Post to K			-0.49*		-0.80**
			(0.21)		(0.31)
Treatment Group Post-Test Level				-0.10	0.44
				(0.15)	(0.30)
Treatment Group Gains Pre-K Post to K				-0.24	0.49
				(0.27)	(0.39)
Constant	0.02	0.62+	0.67*	0.33	0.42
	(0.06)	(0.33)	(0.28)	(0.35)	(0.35)
N (City / Study Blocks)	4 / 31	4 / 31	4 / 31	4 / 31	4 / 31
τ (site)	0.03	0.00	0.00	0.00	0.00
	San Diego Dropped				
	(1)	(2)	(3)	(4)	(5)
Post-Test Impacts (Adj for Pre-Test)	0.46***	0.43**	0.44**	0.53**	0.39+
	(0.14)	(0.15)	(0.14)	(0.17)	(0.23)
Total Post-Test Level		-0.19			
		(0.25)			
Total Gains Between Pre-K Post and K		-0.39			
		(0.49)			
Control Group Post-Test Level			-0.33		-0.56+
			(0.21)		(0.23)
Control Group Gains Pre-K Post to K			-0.95*		-1.24**
			(0.40)		(0.42)
Treatment Group Post-Test Level				0.16	0.54+
				(0.24)	(0.30)
Treatment Group Gains Pre-K Post to K				0.43	0.77+
				(0.38)	(0.40)
Constant	-0.05	0.51	1.35*	-0.65	0.68
	(0.06)	(0.70)	(0.59)	(0.53)	(0.70)
N (City / Study Blocks)	4 / 36	4 / 36	4 / 36	4 / 36	4 / 36
τ (site)	0.00	0.03	0.00	0.00	0.00
	NYC Dropped				
	(1)	(2)	(3)	(4)	(5)
Post-Test Impacts (Adj for Pre-Test)	0.28+	0.34*	0.44*	0.35+	0.41
	(0.15)	(0.16)	(0.17)	(0.19)	(0.25)

Total Post-Test Level		-0.15			
		(0.16)			
Total Gains Between Pre-K Post and K		-0.38			
		(0.38)			
Control Group Post-Test Level			-0.22		-0.57+
			(0.14)		(0.29)
Control Group Gains Pre-K Post to K			-0.52*		-0.91**
			(0.26)		(0.33)
Treatment Group Post-Test Level				0.03	0.46
				(0.16)	(0.31)
Treatment Group Gains Pre-K Post to K				0.31	0.85+
				(0.45)	(0.47)
Constant	0.05	0.52	0.69*	-0.32	0.14
	(0.08)	(0.46)	(0.32)	(0.54)	(0.46)
N (City / Study Blocks)	4 / 25	4 / 25	4 / 25	4 / 25	4 / 25
τ (site)	0.08	0.05	0.00	0.11	0.00

Note. *$p < .05$. **$p < .01$. ***$p < .001$. Standard errors are provided in parentheses. Estimates were derived from a series of regressions that sequentially exclude each study site, such that panel 1 excludes Buffalo, panel 2 excludes Boston, panel 3 excludes Tennessee, panel 4 excludes San Diego, and panel 5 excludes NYC. All models are run at the block level (n=41) and control for children's pre-test assessment. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site and weights.

Table A11

Exploratory Predictors of Follow-Up Impact with Each Study Site Sequentially Dropped

	Buffalo Dropped					
	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts (Adj for Pre-Test)	0.43**	0.36*	0.42**	0.35*	0.36*	0.40**
	(0.16)	(0.16)	(0.15)	(0.17)	(0.14)	(0.15)
Non-Math Cog. Impacts (Pre-K; Adj. for Pre-Test)	-0.25					
	(0.17)					
Impact on Overall COEMET Quality Score (Pre-K)		-0.01				
		(0.05)				
Impact on Math Instructional Time (Pre-K)			-0.06+			
			(0.03)			
Impact on No. of Math Activities (Pre-K)				0.00		
				(0.07)		
Limited English Proficiency (10%)					0.03+	
					(0.02)	
Black-Non Hispanic (10%)						-0.02
						(0.02)
Constant	-0.02	0.00	0.07	-0.00	-0.08	0.08
	(0.07)	(0.07)	(0.06)	(0.08)	(0.07)	(0.07)
N (City / Study Blocks)	4 / 35	4 / 34	4 / 34	4 / 34	4 / 35	4 / 35
τ (site)	0.09	0.08	0.02	0.08	0.00	0.00
	Boston Dropped					
	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts (Adj for Pre-Test)	0.45**	0.42**	0.46***	0.43**	0.43***	0.45***
	(0.15)	(0.14)	(0.13)	(0.14)	(0.12)	(0.13)
Non-Math Cog. Impacts (Pre-K; Adj. for Pre-Test)	-0.18					
	(0.18)					
Impact on Overall COEMET Quality Score (Pre-K)		-0.03				
		(0.04)				
Impact on Math Instructional Time (Pre-K)			-0.08+			
			(0.05)			
Impact on No. of Math Activities (Pre-K)				-0.05		
				(0.05)		
Limited English Proficiency (10%)					0.03	
					(0.02)	
Black-Non Hispanic (10%)						-0.02
						(0.01)
Constant	0.01	0.02	0.09	0.05	-0.05	0.08
	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)
N (City / Study Blocks)	4 / 37	4 / 36	4 / 36	4 / 36	4 / 37	4 / 37
τ (site)	0.07	0.08	0.00	0.05	0.00	0.00

	Tennessee Dropped					
	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts (Adj for Pre-Test)	0.59***	0.55***	0.60***	0.55***	0.57***	0.56***
	(0.15)	(0.14)	(0.14)	(0.15)	(0.14)	(0.15)
Non-Math Cog. Impacts (Pre-K; Adj. for Pre-Test)	-0.15					
	(0.19)					
Impact on Overall COEMET Quality Score (Pre-K)		0.03				
		(0.05)				
Impact on Math Instructional Time (Pre-K)			-0.05			
			(0.03)			
Impact on No. of Math Activities (Pre-K)				0.04		
				(0.06)		
Limited English Proficiency (10%)					0.02	
					(0.02)	
Black-Non Hispanic (10%)						-0.01
						(0.02)
Constant	0.01	0.01	0.07	-0.01	-0.04	0.04
	(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.07)
N (City / Study Blocks)	4 / 31	4 / 30	4 / 30	4 / 30	4 / 31	4 / 31
τ (site)	0.03	0.02	0.00	0.05	0.00	0.00
	San Diego Dropped					
	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts (Adj for Pre-Test)	0.52***	0.49***	0.52***	0.48***	0.47***	0.48**
	(0.15)	(0.14)	(0.14)	(0.14)	(0.13)	(0.15)
Non-Math Cog. Impacts (Pre-K; Adj. for Pre-Test)	-0.24					
	(0.17)					
Impact on Overall COEMET Quality Score (Pre-K)		-0.03				
		(0.05)				
Impact on Math Instructional Time (Pre-K)			-0.05+			
			(0.03)			
Impact on No. of Math Activities (Pre-K)				-0.02		
				(0.05)		
Limited English Proficiency (10%)					0.03	
					(0.05)	
Black-Non Hispanic (10%)						-0.01
						(0.02)
Constant	-0.07	-0.05	0.01	-0.04	-0.08	-0.02
	(0.07)	(0.06)	(0.07)	(0.08)	(0.08)	(0.11)
N (City / Study Blocks)	4 / 36	4 / 35	4 / 35	4 / 35	4 / 36	4 / 36
τ (site)	0.05	0.00	0.00	0.00	0.00	0.00
	NYC Dropped					
	(1)	(2)	(3)	(4)	(5)	(6)

Post-Test Impacts (Adj for Pre-Test)	0.37*	0.29+	0.33*	0.34*	0.33+	0.36*
	(0.18)	(0.17)	(0.15)	(0.17)	(0.18)	(0.16)
Non-Math Cog. Impacts (Pre-K; Adj. for Pre-Test)	-0.17					
	(0.21)					
Impact on Overall COEMET Quality Score (Pre-K)		0.01				
		(0.05)				
Impact on Math Instructional Time (Pre-K)			-0.06*			
			(0.03)			
Impact on No. of Math Activities (Pre-K)				-0.04		
				(0.06)		
Limited English Proficiency (10%)					0.02	
					(0.03)	
Black-Non Hispanic (10%)						-0.02
						(0.02)
Constant	0.02	0.04	0.14+	0.08	-0.00	0.12
	(0.10)	(0.09)	(0.08)	(0.09)	(0.13)	(0.08)
N (City / Study Blocks)	4 / 25	4 / 24	4 / 24	4 / 24	4 / 25	4 / 25
τ (site)	0.10	0.09	0.00	0.06	0.06	0.00

Note. *$p < .05$. **$p < .01$. ***$p < .001$. Standard errors are provided in parentheses. Estimates were derived from a series of regressions that sequentially exclude each study site, such that panel 1 excludes Buffalo, panel 2 excludes Boston, panel 3 excludes Tennessee, panel 4 excludes San Diego, and panel 5 excludes NYC. All models are run at the block level (n=41) and control for children's pre-test assessment. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site and weights. Both Limited English Proficiency and Black Non-Hispanic variables were rescaled to 10 percentage point units.

Table A12

Predictors of Follow-Up Impact Excluding Post-Randomized School Switching in Boston and Buffalo

	(1)	(2)	(3)	(4)	(5)
Post-Test Impacts (Adj for Pre-Test)	0.38**	0.38**	0.42***	0.42**	0.37
	(0.13)	(0.12)	(0.12)	(0.16)	(0.23)
Total Post-Test Level		-0.15			
		(0.14)			
Total Gains Between Pre-K Post and K		-0.39			
		(0.25)			
Control Group Post-Test Level			-0.19		-0.57
			(0.13)		(0.35)
Control Group Gains Between Pre-K Post and K			-0.46*		-0.88**
			(0.20)		(0.32)
Treatment Group Post-Test Level				-0.03	0.46
				(0.15)	(0.38)
Treatment Group Gains Between Pre-K Post and K				0.07	0.66+
				(0.31)	(0.38)
Constant	0.02	0.53	0.63*	-0.05	0.32
	(0.06)	(0.33)	(0.27)	(0.40)	(0.34)
N (City / Study Blocks)	5 / 38	5 / 38	5 / 38	5 / 38	5 / 38
τ (site)	0.06	0.00	0.00	0.08	0.00

Note. *$p < .05$. **$p < .01$. ***$p < .001$. Standard errors are provided in parentheses. All models are run at the block level (n=38) and adjust for pre-test performance. Analyses exclude schools from Boston and Buffalo (n=6) that either switched intervention condition or were added after the random assignment period.

Table A13

Non-Math and Demographic Predictors of Follow-Up Impact Excluding Post-Randomized School Switching in Boston and Buffalo

	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts (Adj for Pre-Test)	0.45**	0.41**	0.46***	0.43**	0.42***	0.45***
	(0.14)	(0.14)	(0.13)	(0.14)	(0.12)	(0.13)
Non-Math Cog. Impacts (Pre-K; Adj. for Pre-Test)	-0.19					
	(0.16)					
Impact on Overall COEMET Quality Score (Pre-K)		-0.03				
		(0.04)				
Impact on Math Instructional Time (Pre-K)			-0.08+			
			(0.05)			
Impact on No. of Math Activities (Pre-K)				-0.04		
				(0.05)		
Limited English Proficiency (10%)					0.03	
					(0.02)	
Black-Non Hispanic (10%)						-0.02
						(0.01)
Constant	0.01	0.03	0.09	0.05	-0.05	0.08
	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)
N (City / Study Blocks)	5 / 38	5 / 37	5 / 37	5 / 37	5 / 38	5 / 38
τ (site)	0.07	0.07	0.00	0.05	0.00	0.00

Note. *p < .05. **p < .01. ***p < .001. Standard errors are provided in parentheses. All models are run at the block level (n=38) and adjust for pre-test performance. Both Limited English Proficiency and Black Non-Hispanic variables were rescaled to 10 percentage point units. Analyses exclude schools from Boston and Buffalo (n=6) that either switched intervention condition or were added after the random assignment period.

Table A14

Math-Related Predictors of Follow-Up Impact Omitting Blocks with Negative Posttest Impacts

	(1)	(2)	(3)	(4)	(5)
Post-Test Impacts (Adj for Pre-Test)	0.34+	0.22	0.20	0.42+	0.26
	(0.19)	(0.23)	(0.21)	(0.24)	(0.28)
Total Post-Test Level		-0.24			
		(0.24)			
Total Gains Between Pre-K Post and K		-0.52			
		(0.43)			
Control Group Post-Test Level			-0.41*		-0.67*
			(0.20)		(0.30)
Control Group Gains Between Pre-K Post and K			-0.78*		-1.01**
			(0.32)		(0.34)
Treatment Group Post-Test Level				0.09	0.45
				(0.24)	(0.31)
Treatment Group Gains Between Pre-K Post and K				0.22	0.61
				(0.38)	(0.39)
Constant	0.03	0.80	1.21*	-0.30	0.64
	(0.10)	(0.64)	(0.49)	(0.56)	(0.63)
N (City / Study Blocks)	5 / 31	5 / 31	5 / 31	5 / 31	5 / 31
τ (site)	0.03	0.05	0.00	0.06	0.00

Note. *$p < .05$. **$p < .01$. ***$p < .001$. Standard errors are provided in parentheses. All models are run at the block level (n=31) and control for children's pre-test assessment. Estimates were generated using a random effects meta-analytic model that included a random effect for study site and weights. Models exclude blocks (n=10) with negative posttest impacts

Table A15

Block-Level Exploratory Predictors of Follow-Up Impact Omitting Negative Impacts

	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts (Adj for Pre-Test)	0.43+	0.38+	0.37+	0.37+	0.33+	0.38+
	(0.23)	(0.23)	(0.19)	(0.20)	(0.20)	(0.20)
Non-Math Cognitive Impacts (Pre-K; Adj. for Pre-Test)	-0.15					
	(0.19)					
Impact on Overall COEMET Quality Score (Pre-K)		-0.01				
		(0.05)				
Impact on Math Instructional Time (Pre-K)			-0.06*			
			(0.03)			
Impact on Number of Math Activities (Pre-K)				-0.02		
				(0.06)		
Limited English Proficiency (10%)					0.02	
					(0.03)	
Black-Non Hispanic (10%)						-0.02
						(0.02)
Constant	0.00	0.02	0.12	0.04	0.01	0.09
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
	5 / 31	5 / 30	5 / 30	5 / 30	5 / 31	5 / 31
τ (site)	0.08	0.05	0.00	0.04	0.04	0.00

Note. *$p < .05$. **$p < .01$. ***$p < .001$. Standard errors are provided in parentheses. All models are run at the block level (n=31) and control for children's pre-test assessment. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site and weights. Both Limited English Proficiency and Black Non-Hispanic variables were rescaled to 10 percentage point units. Models exclude blocks (n=10) with negative posttest impacts.

Table A16

Block-Level Predictors of Follow-Up Impact With Robust Variation

	(1)	(2)	(3)	(4)	(5)
Post-Test Impacts (Adj for Pre-Test)	0.40	0.39	0.44*	0.41	0.36+
	(0.17)	(0.17)	(0.12)	(0.20)	(0.14)
Total Post-Test Level		-0.16+			
		(0.05)			
Total Gains Between Pre-K Post and K		-0.45*			
		(0.11)			
Control Group Post-Test Level			-0.22+		-0.61*
			(0.07)		(0.09)
Control Group Gains Between Pre-K Post and K			-0.53+		-0.91**
			(0.19)		(0.16)
Treatment Group Post-Test Level				-0.01	0.50*
				(0.10)	(0.11)
Treatment Group Gains Between Pre-K Post and K				0.02	0.67*
				(0.32)	(0.18)
Constant	0.00	0.60+	0.71	-0.02	0.34
	(0.07)	(0.14)	(0.27)	(0.44)	(0.18)
N (City / Study Blocks)	5 / 41	5 / 41	5 / 41	5 / 41	5 / 41
τ (site)	0.06	0.00	0.00	0.08	0.00

Note. + $p < 0.1$ *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$. Standard errors are provided in parentheses. All models are run at the block level (n=41) and control for children's pre-test assessment. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site, weights, and robust standard errors.

Table A17

Block-Level Exploratory Predictors of Follow-Up Impact With Robust Variation

	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts (Adj for Pre-test)	0.48+	0.42+	0.47+	0.43+	0.44+	0.46*
	(0.15)	(0.18)	(0.16)	(0.14)	(0.14)	(0.13)
Non-Math Cognitive Impacts (Pre-K; Adj. for Pre-Test)	-0.20+					
	(0.08)					
Impact on Overall COEMET Quality Score (Pre-K)		-0.01				
		(0.04)				
Impact on Math Instructional Time (Pre-K)			-0.06+			
			(0.01)			
Impact on Number of Math Activities (Pre-K)				-0.02		
				(0.05)		
Limited English Proficiency (10%)					0.03+	
					(0.01)	
Black-Non Hispanic (10%)						-0.02
						(0.01)
Constant	-0.01	0.01	0.07	0.02	-0.06	0.07
	(0.06)	(0.06)	(0.06)	(0.08)	(0.04)	(0.07)
N (Site / Study Blocks)	5 / 41	5 / 40	5 / 40	5 / 40	5 / 41	5 / 41
τ (Site)	0.08	0.07	0.00	0.08	0.07	0.00

Note. + $p < 0.1$ *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$. Standard errors are provided in parentheses. All models are run at the block level (n=41) and control for children's pre-test assessment. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site, meta-analytic weights, and robust standard errors. Both Limited English Proficiency and Black Non-Hispanic variables were rescaled to 10 percentage point units.

Table A18

Block-Level Exploratory Predictors of Follow-Up Impact Using Classroom Observation Difference Score Variables

	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts (Adj for Pre-test)	0.47***	0.42**	0.43***	0.42**	0.44***	0.46***
	(0.14)	(0.14)	(0.12)	(0.14)	(0.12)	(0.13)
Non-Math Cognitive Impacts (Pre-K; Adj. for Pre-Test)	-0.20					
	(0.16)					
Difference in Overall COEMET Quality Score (Pre-K)		-0.03				
		(0.18)				
Difference in Math Instructional Time (Pre-K)			-0.53			
			(0.34)			
Difference in Number of Math Activities (Pre-K)				-0.00		
				(0.02)		
Limited English Proficiency (10%)					0.03	
					(0.02)	
Black-Non Hispanic (10%)						-0.02
						(0.01)
Constant	-0.01	0.00	0.08	0.02	-0.06	0.07
	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
N (Site / Study Blocks)	5 / 41	5 / 40	5 / 40	5 / 40	5 / 41	5 / 41
τ (Site)	0.08	0.07	0.03	0.07	0.00	0.00

Note. + $p < 0.1$ *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$. Standard errors are provided in parentheses. All models are run at the block level (n=41) and control for children's pre-test assessment. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site and meta-analytic weights. The "Difference" variables were generated by taking the block level average of treatment and control scores on each COEMET related measure (i.e., COEMET quality, math instructional time, number of math activities), and using the difference between these two groups as a predictor in each model. Both Limited English Proficiency and Black Non-Hispanic variables were rescaled to 10 percentage point units.

Table A19

Descriptive Statistics on Non-Math Measures by Study Site (Raw Values)

	Buffalo		**Boston**		**Tennessee**		**San Diego**		**NYC**	
	M	SD	M	SD	M	SD	M	SD	M	SD
Pre-K Literacy Score	0.49	0.23	0.77	0.30	-	-	-	-	-	-
Woodcock Johnson -III Letter Word	-	-	-	-	348.37	22.46	-	-	-	-
Renfrew Bus Story	0.01	0.83	0.03	0.87	<0.01	0.83	47.83	9.14	-	-
Head Toes Knees Shoulders (HTKS)	-	-	-	-	-	-	15.90	14.28	-	-
Peg Tapping	-	-	-	-	-	-	10.30	5.98	-	-
Forward Digit Span	-	-	-	-	-	-	3.82	1.30	-	-
Backward Digit Span	-	-	-	-	-	-	0.54	1.03	-	-
Item Selection	-	-	-	-	-	-	11.13	3.17	-	-
Self-Ordered Pointing	-	-	-	-	-	-	17.13	3.42	-	-
PALS- Name Writing	-	-	-	-	-	-	5.91	1.68	-	-
PALS- Alphabet Knowledge	-	-	-	-	-	-	38.91	27.89	-	-
Expressive Vocabulary Test	-	-	-	-	-	-	54.67	20.43	-	-
Receptive One-Word Picture Vocabulary Test (ROWPVT)	-	-	-	-	-	-	-	-	96.51	15.49
Spatial Conflict Arrows	-	-	-	-	-	-	-	-	0.68	0.26
Corsi Blocks Assessment Forwards Score	-	-	-	-	-	-	-	-	3.06	1.04
Corsi Blocks Assessment Backwards Score	-	-	-	-	-	-	-	-	1.33	1.26
Attention and Impulse Control (PSRA)	-	-	-	-	-	-	-	-	2.60	0.47
Non-Math Composite	-0.03	0.86	0.03	0.97	<0.01	0.81	0.01	0.61	-0.01	0.61
Standardized Non-Math Composite	-0.04	0.99	0.03	1.05	-0.00	1.00	0.02	1.01	-0.01	1.01
Observations		510		153		686		524		1216

Note. Non-math assessments are provided for the end of pre-k using *all* available data for each study site. To generate non-math composite scores, all available non-math assessments (z-scored) were averaged together and re-standardized to reduce any bias in the availability of non-math measures and sub-score data. For Buffalo and Boston, the Pre-K Literacy score reflects composite scores of the mCLASS:CIRCLE and PALS Pre-K.

Table A20

Within-Block Average Number of Observations and Standard Errors

	Mean	SD	Min	Max
Post-Test N	76.51	44.42	23.00	291.00
Follow-Up N	70.29	39.26	22.00	251.00
Post-Test SE (No Pre-test adj.)	0.27	0.10	0.11	0.61
Post-Test SE (Adj. for pre-test)	0.21	0.06	0.08	0.37
Follow-Up SE (No Pre-test adj.)	0.29	0.11	0.12	0.74
Follow-Up SE (Adj. for pre-test)	0.24	0.08	0.10	0.51

Table A21

Block-Level Predictors of Follow-Up Impact Controlling for Pre-Test Impacts

	(1)	(2)	(3)	(4)	(5)	(6)
Pre-Test Impacts	-0.05	0.12	0.06	0.05	0.14	0.02
	(0.14)	(0.15)	(0.16)	(0.16)	(0.16)	(0.26)
Post-Test Impacts (Adj for Pre-Test)		0.44***	0.41**	0.46***	0.48**	0.38
		(0.14)	(0.13)	(0.13)	(0.17)	(0.37)
Total Post-Test Level			-0.15			
			(0.15)			
Total Gains Between Pre-K Post and K			-0.43+			
			(0.25)			
Control Group Post-Test Level				-0.20		-0.59
				(0.14)		(0.39)
Control Group Gains Between Pre-K Post and K				-0.52**		-0.91***
				(0.20)		(0.27)
Treatment Group Post-Test Level					0.02	0.48
					(0.15)	(0.37)
Treatment Group Gains Between Pre-K Post and K					0.12	0.68+
					(0.32)	(0.36)
Constant	0.10 (0.07)	-0.02	0.56 (0.34)	0.69*	-0.18	0.32 (0.40)
		(0.07)		(0.28)	(0.43)	
N (City / Study Blocks)	5 / 41	5 / 41	5 / 41	5 / 41	5 / 41	5 / 41
τ (site)	0.10	0.06	0.00	0.00	0.09	0.00

Note. + $p < 0.1$ *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$. Standard errors are provided in parentheses. All models are run at the block level (n=41) and control for children's pre-test assessment. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site and weights. The tau represents between study variation in effects (interpreted in standard deviation units).

Table A22

Block-Level Exploratory Predictors of Follow-Up Impact Controlling for Pre-Test Impacts

	(1)	(2)	(3)	(4)	(5)	(6)
Pre-Test Impacts	0.09	0.15	0.11	0.12	0.03	0.09
	(0.16)	(0.17)	(0.15)	(0.16)	(0.17)	(0.15)
Post-Test Impacts (Adj for Pre-test)	0.50***	0.48***	0.50***	0.47***	0.45***	0.49***
	(0.15)	(0.15)	(0.13)	(0.15)	(0.13)	(0.14)
Non-Math Cognitive Impacts (Pre-K; Adj. for Pre-Test)	-0.19					
	(0.16)					
Impact on Overall COEMET Quality Score (Pre-K)		-0.03				
		(0.05)				
Impact on Math Instructional Time (Pre-K)			-0.06*			
			(0.03)			
Impact on Number of Math Activities (Pre-K)				-0.03		
				(0.05)		
Limited English Proficiency (10%)					0.03	
					(0.02)	
Black-Non Hispanic (10%)						-0.02
						(0.01)
Constant	-0.03	-0.02	0.05	0.00	-0.06	0.05
	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)
N (Site / Study Blocks)	5 / 41	5 / 40	5 / 40	5 / 40	5 / 41	5 / 41
τ (Site)	0.08	0.06	0.00	0.06	0.00	0.00

Note. + $p < 0.1$ *$p < 0.05$. **$p < 0.01$. ***p < 0.001. Standard errors are provided in parentheses. All models are run at the block level (n=41) and control for children's pre-test assessment. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site and meta-analytic weights. The tau represents between study variation in effects (interpreted in standard deviation units). Both Limited English Proficiency and Black Non-Hispanic variables were rescaled to 10 percentage point units.

Table A23

Block-Level Predictors of Follow-Up Impact Excluding Blocks with Over 10% Attrition

	(1)	(2)	(3)	(4)	(5)
Post-Test Impacts (Adj for Pre-Test)	0.34*	0.32+	0.32+	0.48*	0.38
	(0.17)	(0.18)	(0.19)	(0.19)	(0.29)
Total Post-Test Level		-0.18			
		(0.25)			
Total Gains Between Pre-K Post and K		0.02			
		(0.54)			
Control Group Post-Test Level			-0.25		-0.49
			(0.23)		(0.46)
Control Group Gains Between Pre-K Post and K			-0.69		-1.11+
			(0.48)		(0.60)
Treatment Group Post-Test Level				-0.07	0.38
				(0.23)	(0.45)
Treatment Group Gains Between Pre-K Post and K				0.41	0.71+
				(0.38)	(0.40)
Constant	0.05	0.06	1.04	-0.45	0.65
	(0.10)	(0.75)	(0.69)	(0.52)	(0.80)
N (City / Study Blocks)	5 / 26	5 / 26	5 / 26	5 / 26	5 / 26
τ (site)	0.09	0.09	0.07	0.06	0.00

Note. + $p < 0.1$ *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$. Standard errors are provided in parentheses. All models are run at the block level (n=41) and control for children's pre-test assessment. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site and weights. The tau represents between study variation in effects (interpreted in standard deviation units).

Table A24

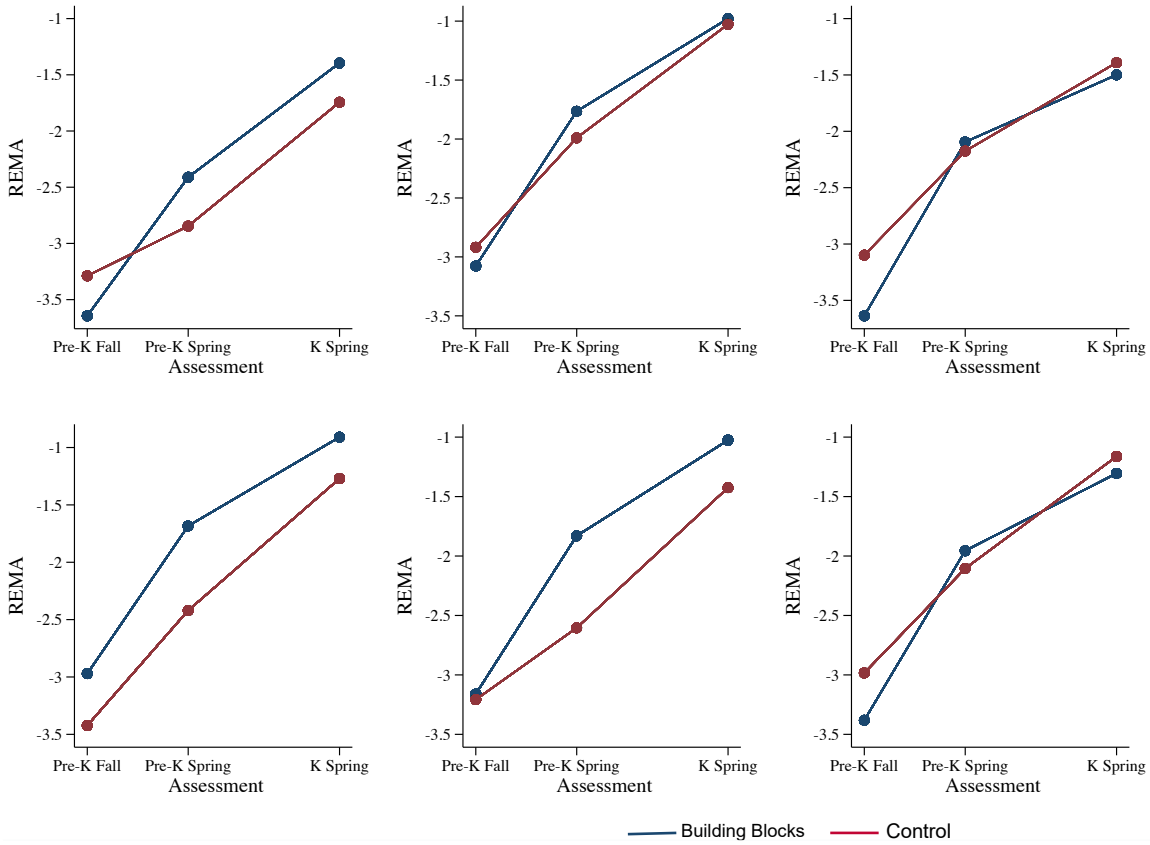
Block-Level Exploratory Predictors of Follow-Up Impact Excluding Blocks with Over 10% Attrition

	(1)	(2)	(3)	(4)	(5)	(6)
Post-Test Impacts (Adj for Pre-test)	0.36+	0.38*	0.36*	0.36*	0.34*	0.35*
	(0.20)	(0.18)	(0.15)	(0.16)	(0.17)	(0.16)
Non-Math Cognitive Impacts (Pre-K; Adj. for Pre-Test)	-0.04					
	(0.23)					
Impact on Overall COEMET Quality Score (Pre-K)		-0.03				
		(0.05)				
Impact on Math Instructional Time (Pre-K)			-0.08+			
			(0.05)			
Impact on Number of Math Activities (Pre-K)				-0.07		
				(0.06)		
Limited English Proficiency (10%)					0.01	
					(0.04)	
Black-Non Hispanic (10%)						-0.02
						(0.02)
Constant	0.04	0.05	0.14	0.12	0.03	0.13
	(0.10)	(0.09)	(0.10)	(0.11)	(0.11)	(0.13)
N (Site / Study Blocks)	5 / 26	5 / 26	5 / 26	5 / 26	5 / 26	5 / 26
τ (Site)	0.09	0.08	0.00	0.07	0.10	0.05

Note. + $p < 0.1$ *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$. Standard errors are provided in parentheses. All models are run at the block level (n=41) and control for children's pre-test assessment. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study site and meta-analytic weights. The tau represents between study variation in effects (interpreted in standard deviation units). Both Limited English Proficiency and Black Non-Hispanic variables were rescaled to 10 percentage point units.

Figure A1

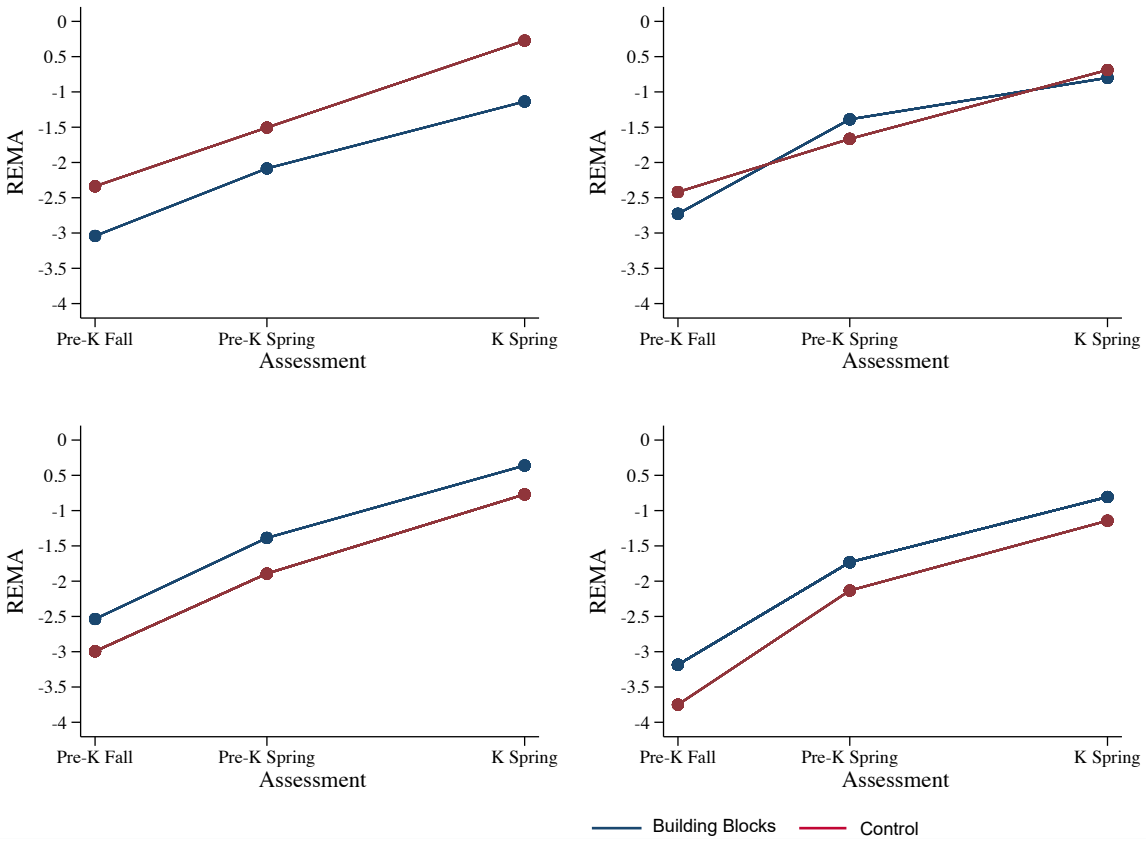
Buffalo REMA Block-Level Trajectories



Note. This graph portrays control and treatment group REMA scores at pre-k fall (pre-test), pre-k spring (post-test), and kindergarten ("K") spring (1-year follow-up) within each block for the Buffalo site.

Figure A2

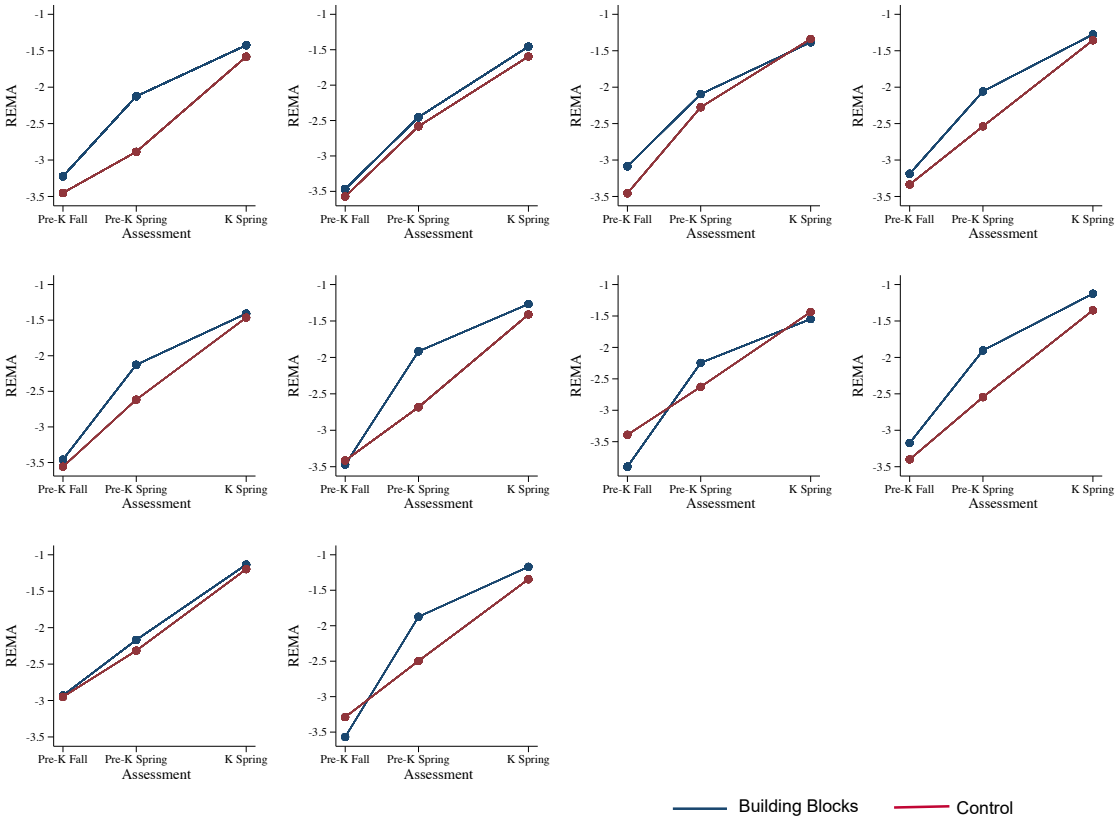
Boston REMA Block-Level Trajectories



Note. This graph portrays control and treatment group REMA scores at pre-k fall (pre-test), pre-k spring (post-test), and kindergarten ("K") spring (1-year follow-up) within each block for the Boston site.

Figure A3

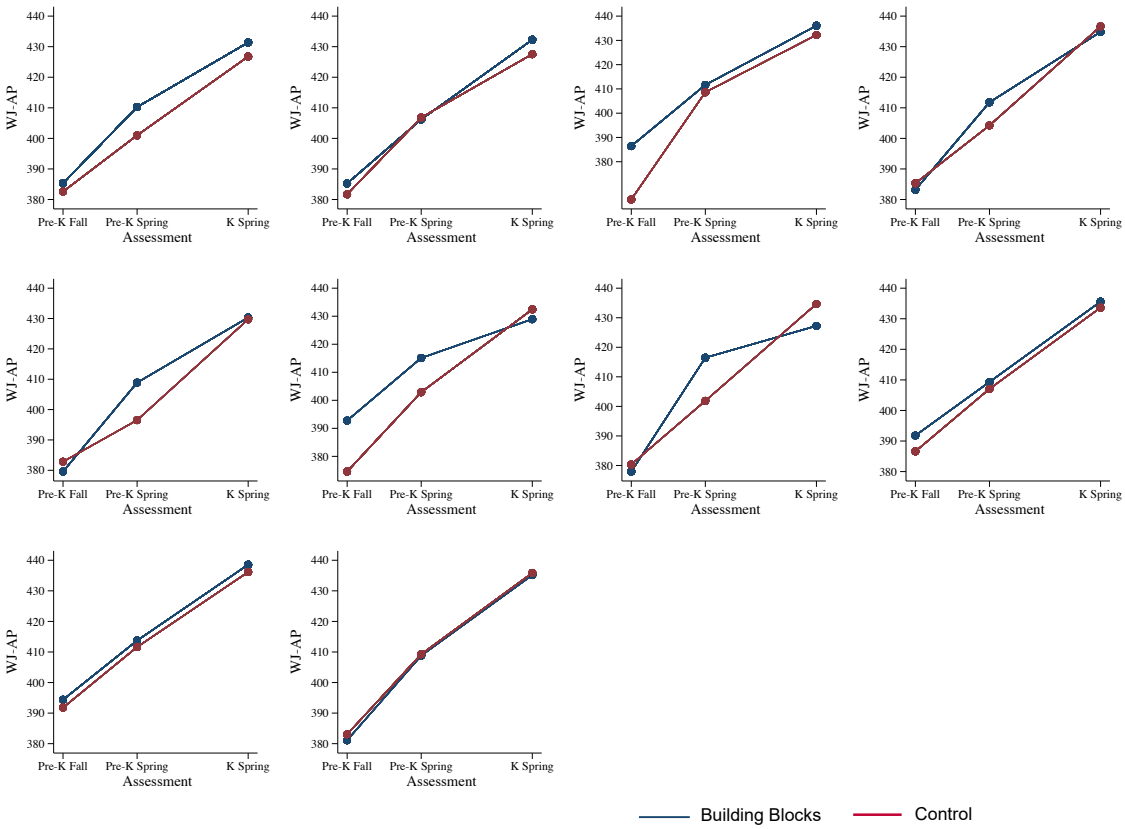
Tennessee REMA Block-Level Trajectories



Note. This graph portrays control and treatment group REMA scores at pre-k fall (pre-test), pre-k spring (post-test), and kindergarten ("K") spring (1-year follow-up) within each block for the Tennessee site.

Figure A4

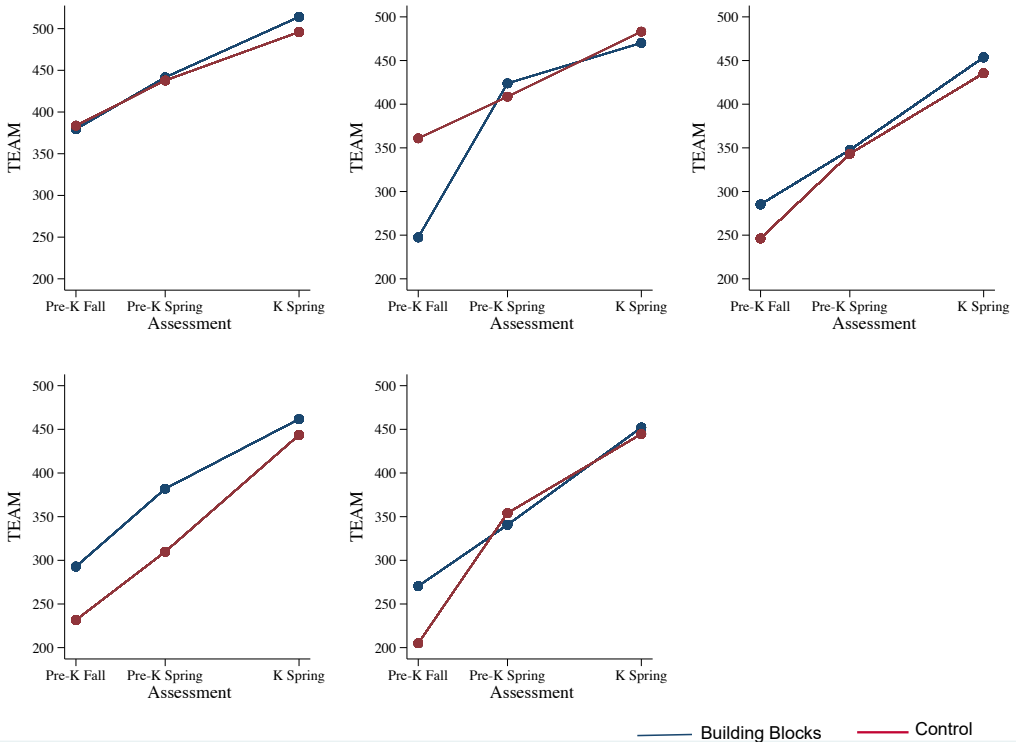
Tennessee Woodcock-Johnson Applied Problems Block-Level Trajectories



Note. This graph portrays control and treatment group Woodcock Johnson III – Applied Problems scores at pre-k fall (pre-test), pre-k spring (post-test), and kindergarten ("K") spring (1-year follow-up) within each block for the Tennessee site.

Figure A5

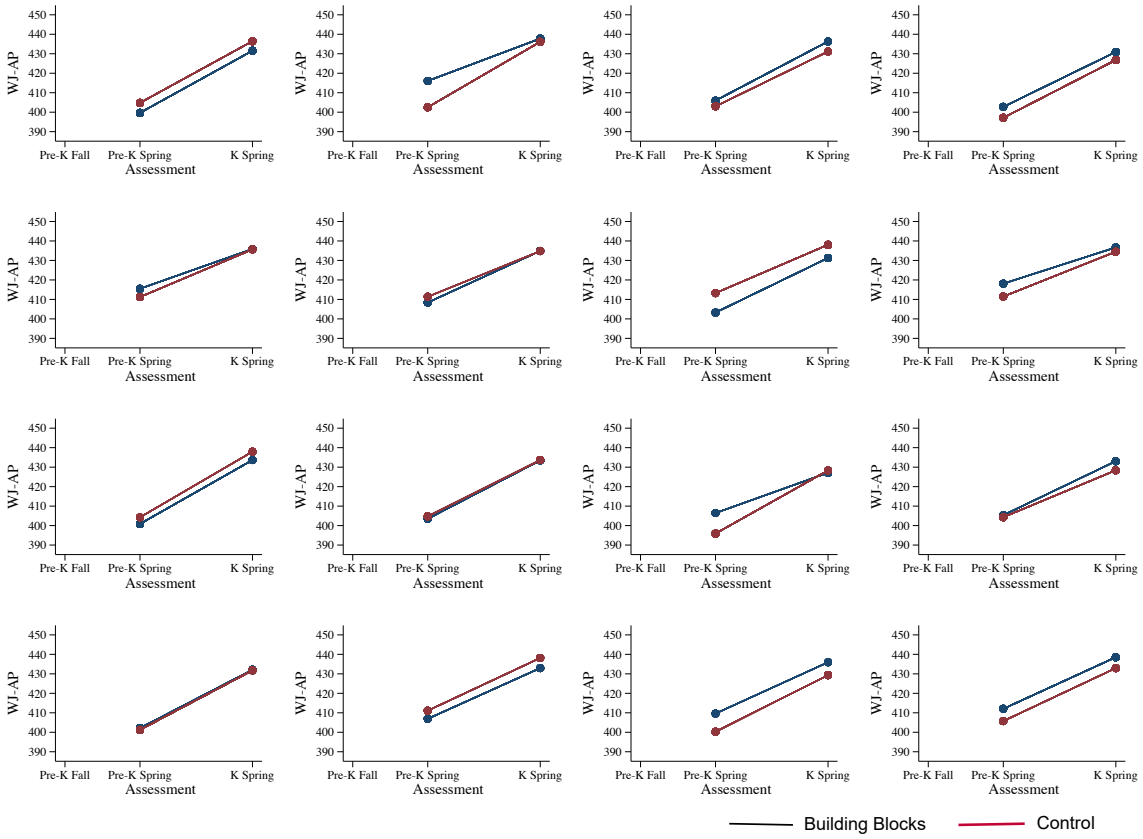
San Diego TEAM Block-Level Trajectories



Note. This graph portrays control and treatment group TEAM scores at pre-k fall (pre-test), pre-k spring (post-test), and kindergarten ("K") spring (1-year follow-up) within each block for the San Diego site.

Figure A6

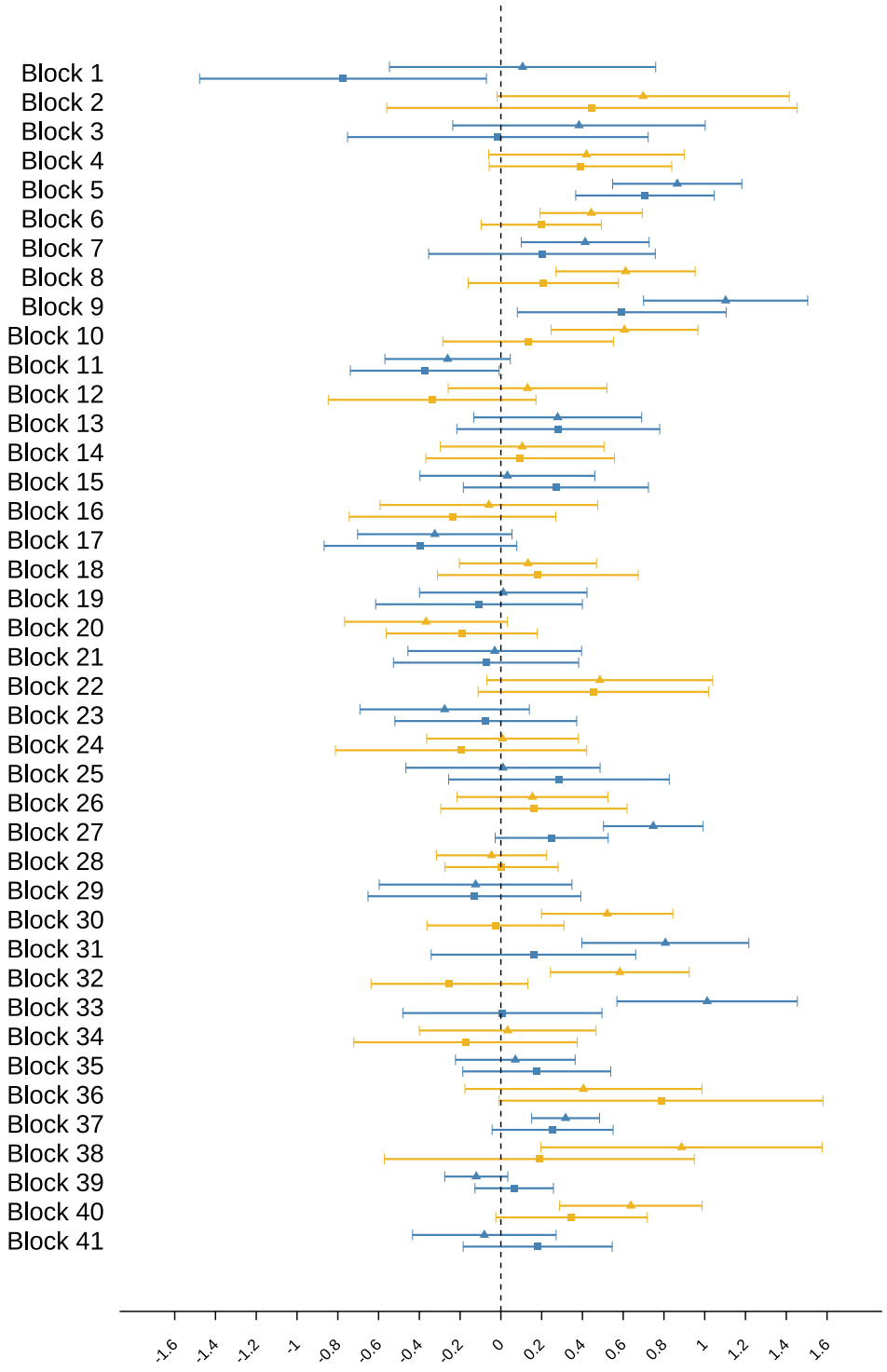
NYC Woodcock-Johnson Applied Problems Block-Level Trajectories



Note. This graph portrays control and treatment group Woodcock Johnson III – Applied Problems scores at pre-k fall (pre-test), pre-k spring (post-test), and kindergarten ("K") spring (1-year follow-up) within each block for the NYC site.

Figure A7

Forest Plot of Post-Test and Follow-up Effect Sizes for Each Block



Note. Triangle = post-test effect size; Square = 1-year follow-up effect size.