

Results of the Early Math Project – Scale-Up Cross-Site Results

Working Paper

Kerry G. Hofer, Ph.D.
Mark W. Lipsey, Ph.D.
Nianbo Dong, Ph.D.
Dale C. Farran, Ph.D.

June 2013

Director

Mark W. Lipsey, Ph. D.

Senior Associate Director

Dale C. Farran, Ph. D.

Associate Director

Sandra Jo Wilson, Ph. D.

Peabody Research Institute

Vanderbilt University

Our mission is to conduct research aimed at improving the effectiveness of programs for children, youth, and families. Using field research, program evaluation, and research synthesis (meta-analysis), our faculty and staff help determine which programs are actually making a difference in the lives of the people they serve. PRI research addresses many aspects of child and family programs, such as their implementation, costs, dissemination, and social or political support. But the main focus for all of our work is the effects of programs on children and families.

Recommended Citation:

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)*. Nashville, TN: Vanderbilt University, Peabody Research Institute.

Funding Source:

This research was conducted in partnership with Doug Clements and Julie Sarama at the University at Buffalo SUNY (recently moved to the University of Colorado). This report includes data from the three sites of the project – Tennessee, New York, and Massachusetts. The latter two sites contributed their data for this report through Clements and Sarama, and all data were synthesized and analyzed at the Tennessee site.

The project was funded through the Institute of Education Sciences and was awarded to Principal Investigators Doug Clements and Julie Sarama (Technology-enhanced Research-based Instruction, Assessment, and Professional Development, #R305K050157), with a subcontract to Dale Farran and Mark Lipsey, Vanderbilt University.

Contact Us:

Phone: 615.322.8540

Fax: 615.322.0293

Mailing Address:

Delivery Address:

Peabody Research Institute

Peabody Research Institute

230 Appleton Place

1930 South Drive

PMB 181

Room 410A

Nashville, TN 37203-5721

Nashville, TN 37212

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

Disclaimer

The opinions expressed in this report are the authors' and do not necessarily represent the opinions and positions of the Institute of Education Sciences of the U. S. Department of Education.

TRIAD Project Tennessee Staff, 2006-2012

Principal Investigator: Dale C. Farran, Senior Associate Director, Peabody Research Institute and Professor, Dept. of Teaching and Learning;
Dale.Farran@vanderbilt.edu.

Co-Principal Investigator: Mark W. Lipsey, Director, Peabody Research Institute;
Mark.Lipsey@vanderbilt.edu.

Project Coordinators: Carol Bilbrey and Rachael Tanner-Smith.

Data Managers/Analysts: Kerry G. Hofer, Nianbo Dong, and Sandra Wilson.

Research Assistants: Beth Vorhaus, Ilknur Sekmen, Christy Plummer, Catherine Fleming, and Marianne Reale.

Graduate Research Assistants: Karen Anthony, Canan Aydogan, Tracy Cummings, Linda Dake, Kelley Drennan, Sarah Shufelt, Filiz Varol, and Betsy Watson.

Facilitators: Susan Brown (Metro Action Council), Anita Conn (Metro Action Council), Penny Franklin (Metropolitan Nashville Public Schools), and Patty Turner (Metropolitan Nashville Public Schools).

Assessors and Observers across Middle Tennessee.

Results of the Early Math Project – Scale-Up Cross-Site Results

Table of Contents	Page
Staff and Contact Information.....	2
Table of Contents.....	3
List of Tables.....	4
List of Figures.....	7
Summary	8
Study Design	8
Intervention Description	8
Professional Development and Coaching.....	9
Sample.....	9
Children.....	10
Measures	10
Direct Assessments	10
Classroom and Child Observation Measures.....	10
Data Analysis and Results.....	11
Descriptives.....	11
Missing Data	11
Baseline Equivalence.....	11
Research Question 1: Immediate and Longitudinal Intervention Effects	15
Research Question 2: Variation in Effects across Sites.....	15
Research Question 3: Curricular Effects for Subgroups of Children.....	18
Research Question4: Effects of Math Environment and Fidelity of Implementation on Gain.....	19
Summary of Results by Research Question.....	27
References	29
Appendices.....	30
Appendix A.....	31
Appendix B.....	73

List of Tables

- 1: Sample of Schools and Classrooms by Site and Condition
- 2: REMA T-Score Descriptive Statistics (Unadjusted, Unimputed) by Condition and Site, Complete Sample
- 3: Complete and Common Samples: Baseline Covariate Comparison across Conditions
- 4: Complete and Common Samples: Baseline REMA Pretest Randomization Check (Overall and by Site)
- 5: Complete and Common Samples: Main Effects Results for Each Time Point
- 6: Individual Site Contrasts from the Model Including the Site x Condition Interaction (Complete Sample)
- 7: Condition Effects on REMA Scores for Sites Analyzed Separately
- 8: Condition x Child Characteristic Interactions in the Complete Sample (each moderator examined in a separate analysis)
- 9: Descriptive Statistics for Pre-K Classrooms on the COEMET by Condition and Site
- 10: Descriptive Statistics for Treatment Classrooms on the Near Fidelity Subscales by Site
- 11: Near Fidelity Additive Score Descriptives by Site and Overall
- 12: Predicting Pre-K gain from Separate Near Fidelity Setting Factor Scores
- 13: Correlations between the COEMET Composite and Near Fidelity Composite across Treatment Classrooms (Overall and by Site)

- A1: Sample Sizes by Time (Series 1 – complete sample)
- A2: Descriptive Statistics by Condition: Child Level Variables (Series 1)
- A3: Descriptive Statistics by Site and Condition: Child Level Variables (Series 1)
- A4: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample: Time 2 (Series 1)
- A5: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample: Time 3 (Series 1)
- A6: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample: Time 4 (Series 1)
- A7: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample with Condition x Site Interaction: Time 2 (Series 1)
- A8: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample with Condition x Site Interaction: Time 3 (Series 1)
- A9: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample with Condition x Site Interaction: Time 4 (Series 1)
- A10: Fixed and Random Effects Analysis for REMA Outcome -By site: Time 2 (Series 1)
- A11: Fixed and Random Effects Analysis for REMA Outcome -By site: Time 3 (Series 1)
- A12: Fixed and Random Effects Analysis for REMA Outcome -By site: Time 4 (Series 1)
- A13: Summary of Moderator Analyses for REMA Outcome: Time 2 (Series 1)
- A14: Summary of Moderator Analyses for REMA Outcome: Time 3 (Series 1)
- A15: Summary of Moderator Analyses for REMA Outcome: Time 4 (Series 1)
- A16: 3-way Interaction of Pretest by Condition by Site for REMA Outcome (Complete Sample)
- A17: Pretest (site-mean centering) as a Moderator by Site for REMA Outcome (Complete Sample: Time 2)

- A18: Pretest (site-mean centering) as a Moderator by Site for REMA Outcome (Complete Sample: Time 3)
- A19: Pretest (site-mean centering) as a Moderator by Site for REMA Outcome (Complete Sample: Time 4)
- A20: P-values of Testing Differences of COEMET Measures across Site and Condition
- A21: Detailed Results of Testing Differences of COEMET Measures across Site and Condition
- A22: Factor Loadings of the COEMET Composite Variable
- A23: COEMET Composite Variable Predicting Pre-K REMA Gain (Series 1: Time 2)
- A24: COEMET Site-Centered Subscales Predicting REMA Pre-K Gain (Series 1: Time 2)
- A25: COEMET Composite Variable Predicting REMA Gain through K (Series 1: Time 3)
- A26: COEMET Site-Centered Subscales Predicting REMA Gain through K (Series 1: Time 3)
- A27: COEMET Composite Variable Predicting REMA Gain through 1st Grade (Series 1: Time 4)
- A28: COEMET Site-Centered Subscales Predicting REMA Gain through 1st Grade (Series 1: Time 4)
- A29: The COEMET Composite Variable as a Moderator of the Effects of Condition on REMA Gain During Pre-K (Series 1: Time 2)
- A30: Analysis of the COEMET Composite Variable as a Mediator the Effects of Condition on the REMA During Pre-K. Part 1: Sample Size and Model Fit Statistics (Series 1: Time 2)
- A31: Analysis of the COEMET Composite Variable as a Mediator the Effects of Condition on REMA During Pre-K. Part 2: Level 1 Within Classroom Results (Series 1: Time 2)
- A32: Analysis of the COEMET Composite Variable as a Mediator the Effects of Condition on REMA During Pre-K. Part 3: Level 2 Between School Results (Series 1: Time 2)
- A33: Factor Loadings of the Near Fidelity Items Within Setting
- A34: Zero-order Correlations among Near Fidelity Scores and Classroom Mean Residualized Gain
- A35: Fixed and Random Effects of Near Fidelity Composite: Time 2
- A36: Fixed Effects of Near Fidelity Setting Variables: Time 2 (Full Treatment Sample)
- A37: Fixed Effects of Near Fidelity Setting Variables: Time 2 (Reduced Sample)
- A38: Fixed Effects of Near Fidelity Additive Composite on REMA Gain, controlling for COEMET Composite: Time 2

- B1: Sample Sizes by Time (Series 2 – Common Sample)
- B2: Descriptive Statistics by Condition across time: Child Level Variables (Series 2)
- B3: Descriptive Statistics by Site and Condition: Child Level Variables (Series 2)
- B4: Fixed and Random Effects for the REMA Outcome –Common Sample: Time 2 (Series 2)
- B5: Fixed and Random Effects for the REMA Outcome -Common Sample: Time 3 (Series 2)
- B6: Fixed and Random Effects for the REMA Outcome -Common Sample: Time 4 (Series 2)
- B7: Fixed and Random Effects for the REMA Outcome –Common Sample with Condition x Site Interaction: Time 2 (Series 2)

- B8: Fixed and Random Effects for the REMA Outcome -Common Sample with Condition x Site Interaction: Time 3 (Series 2)
- B9: Fixed and Random Effects for the REMA Outcome -Common Sample with Condition x Site Interaction: Time 4 (Series 2)
- B10: Fixed and Random Effects for the REMA Outcome -By site: Time 2 (Series 2)
- B11: Fixed and Random Effects for the REMA Outcome -By site: Time 3 (Series 2)
- B12: Fixed and Random Effects for the REMA Outcome -By site: Time 4 (Series 2)
- B13: Summary of Moderator Analyses for REMA Outcome: Time 2 (Series 2)
- B14: Summary of Moderator Analyses for REMA Outcome: Time 3 (Series 2)
- B15: Summary of Moderator Analyses for REMA Outcome: Time 4 (Series 2)

List of Figures

- 1: Covariate-adjusted REMA t-scores across time by setting and condition
 2. Level 1 of SEM Mediation Analysis: Classrooms within Schools
 3. Level 2 of SEM Mediation Analysis: Between Schools
-
- A1: Least Square REMA Means by Condition and Time (Series 1)
 - A2: Effect Size on the REMA by Condition and Time (Series 1)
 - A3: Least Square REMA Means by Site, Condition, and Time (Series 1)
-
- B1: Least Square REMA Means by Condition and Time (Series 2)
 - B2: Effect Size on the REMA by Condition and Time (Series 2)
 - B3: Least Square REMA Means by Site, Condition, and Time (Series 2)

Results of the Early Math Project – Scale-Up Cross-Site Results

Summary

This report summarizes the results of a scale-up project funded by the Institute of Education Sciences in 2006. *Scaling up TRIAD: Teaching Early Mathematics for Understanding with Trajectories and Technologies* was a project that took a preschool mathematics intervention to scale across three sites, following children from their Prekindergarten year -- in which the intervention was delivered -- through first grade.

This study was designed to explore the following research questions:

1. What are the immediate and long-term effects of the intervention on children’s math skills?
2. How much variation was there in effects across sites?
3. Were curricular effects different for different subgroups of children?
4. What are the effects of the math environment and the fidelity of implementation on children’s immediate and long term math gains, overall and across sites?

Study Design

In order to answer the research questions of interest, we conducted a cluster randomized controlled trial in which schools were randomly assigned either to participate in the treatment condition or in the counterfactual condition (business as usual). This scale-up intervention took place in preschool classrooms in three urban school districts: two Northern city school systems (Boston, MA and Buffalo, NY) and a combination of a metropolitan public school system (referred to from this point as Metro) and a Head Start program (hereafter referred to as Head Start) in Nashville, TN. This report includes only public school programs, as those were common across all three sites. The original site where the curriculum was developed and the developers lived was Buffalo, NY. Boston and Nashville were the distant sites, with Nashville being the most independent.

Intervention Description. This project was designed to evaluate the cross-site execution of Technology-enhanced Research-based Instruction, Assessment, and Professional Development (TRIAD), a model for the implementation of an early mathematics curriculum (Clements & Sarama, 2006). Generally, TRIAD focuses on the math environments that the children are exposed to at home and in the classroom, with a comprehensive professional development package for teachers. The intervention involved the training of Pre-K teachers in their own early math knowledge and in the curricular components, the coaching of those teachers throughout the study, the implementation of the curriculum in real-life classrooms, the supplying of the classrooms with needed materials, and the evaluation of the classroom, teachers, and students involved. The curriculum itself, *Building Blocks*, approaches children’s mathematical learning as a series of research-based learning trajectories. The program attempts to facilitate children’s growth in those trajectories by helping them “mathematize their everyday activities, from building blocks

to art and stories to puzzles and games” (Clements & Sarama, 2011, p. 969). A comprehensive curriculum, *Building Blocks* incorporates 30 weekly lesson plans through small group structures, computer activities, center-based learning, whole group instruction, family letters, and teacher tracking of children’s progress along the math trajectories.

In the Massachusetts and New York sites, children were initially randomized into one of three conditions, with the third unique condition being a *Building Blocks* follow-through group (BBFT). This group received the same Pre-K intervention as the one-year-only *Building Blocks* condition. However, they received additional *Building Blocks* instruction in Kindergarten and 1st grade. Because this condition was not used in the Tennessee site, these children were only included in the analyses involving Pre-K year effects, but this is described in greater detail further in this report.

Professional Development and Coaching. The professional development of teachers in the experimental condition included workshops, in-classroom mentoring, and continued support through an online resource. All *Building Blocks* teachers (unless they came to the project late) participated in three Building Blocks Workshops held throughout the year, 4 days in August of 2006, 2 days in January 2007, and 1 day in April of 2007.

Sample

The scale-up study involved 139 public Pre-K classrooms, described in number by site and condition in Table 1. The intervention in the Tennessee site lagged the other two sites by one year. However, the same protocol was followed in all three sites: one year of teacher training followed by full implementation of the intervention in the second year in Pre-K classrooms, with an additional follow-up of children as they progressed into Kindergarten and first grade. Teachers in the control classrooms practiced business-as-usual instruction and were promised training in the *Building Blocks* curriculum after the evaluation ended if their school systems wished them to receive it.

Table 1. Sample of Schools and Classrooms by Site and Condition

	Classrooms	Schools
Buffalo	75	25
Treatment	51	15
Control	24	10
Boston	31	18
Treatment	21	12
Control	10	6
Nashville	33	16
Treatment	16	8
Control	17	8

Children. Data used in these cross-site analyses were limited to public Pre-K from TN (no Head Start cases) so that, for consistency, only public Pre-K children were represented in all the analyses. Additionally, BBFT children in Buffalo were only used for analyses involving Pre-K effects but were not included in analyses after the point where they participated a different experimental condition, i.e., in Kindergarten and 1st grade. We began with a dataset of 1828 children, 1827 of whom had valid pretest scores.

The **Complete Sample** was defined as those children who had posttest data at the time point to be analyzed. This sample was analyzed at Times 2 (end of Pre-K), 3 (end of K), and 4 (end of 1st grade). The sample size varied across Time — at the end of Pre-K 1714 children in 139 classrooms in 59 schools were included; at Kindergarten and 1st grade, 1192 and 1129 children were included respectively from 103 of the original Pre-K classrooms in 46 schools. (The children were in many more K and 1st grade classrooms.) The numbers in the Complete Sample, broken out by site, can be found in Table A1 in Appendix A.

A second **Common Sample** was defined as those children who were not in the BBFT condition and had valid outcome data at each time point. This sample allowed effects to be examined across the time points with the same children represented at each time. This Common Sample included 1105 children from 103 Pre-K classrooms in 46 schools; the numbers in this sample are broken out by site in Table B1 in Appendix B

Measures

Direct Assessments. All children in the cross-site study were directly assessed with a non-standardized measure described in greater detail below. Children were individually assessed outside the classroom by trained assessors in a quiet location within the school. Tests were administered at the beginning of Pre-K, at the end of Pre-K, at the end of Kindergarten, and at the end of first grade.

- *Research-based Elementary Math Assessment (REMA).* The REMA (Clements, Sarama, & Liu, 2008) is a proximal measure of children’s early math skills, one that is closely aligned with the *Building Blocks* curriculum and designed by the curriculum developers. It includes both number sense and geometry/measurement components that were combined through Rasch modeling to yield one total score (Clements, Sarama, & Liu). The REMA was originally developed for Prekindergarten but items were added to extend its suitability to Kindergarten and first grade. The REMA assesses a child’s developmental progression in skills like verbal counting, subitizing, number comparison, number composition, shape recognition, patterning, spatial imagery, geometric measurement, etc.

Classroom and Child Observation Measures. Each participating Pre-K classroom was observed by trained and reliable project staff using multiple instruments. Classrooms were observed two or three different times during the Pre-K year, depending on site (near the beginning of the year, near the middle of the year, and near the end of the year). Observers used two measures of instruction – one focused on general math instruction used in both treatment and comparison classrooms and one specific to the curriculum used only in

treatment classrooms. These instruments are described in greater detail below. Final scores on all variables of interest, unless otherwise indicated, were averaged across observation periods yielding a more stable measure.

- *Classroom Observation of Early Mathematics – Environment and Teaching (COEMET)*. Used in both treatment and control classrooms, the COEMET (Sarama & Clements, 2007) is an instrument that assesses the instructional environment of the classroom, focusing on the math content specifically. It is comprised of several parts. The Classroom Culture portion, which ranges from 1-5 (5 being the best), the general math environment throughout the observation; it includes questions about how actively the teacher interacts with the children, how the teacher uses teachable math moments, how math is displayed in the physical environmental of the room, how confident the teacher appeared about math, etc.

Each different instance of math instruction conducted by the teacher during the observation is cataloged as a Specific Math Activity, or SMA. The SMA portion of the COEMET uses event sampling; each time a math activity is observed, questions about its length, instructional quality, teacher involvement, child engagement, and other characteristics are answered by the observer, yielding a count for the number of observed instances, an overall quality score averaged across incidents (also ranging from 1-5), and the time length for each SMA. An example of an SMA might be when a teacher engages the whole class as part of the morning meeting time in a guided activity that is focused on math (counting students present, discussing the days in school and days left until the weekend, using the calendar to talk about number composition, etc.). A teaching incident had to occur for at least a minute to be considered an SMA.

As often happens in a Pre-K classroom, a teacher might organize a small group activity and rotate students through the activity in groups of 3-4. When this occurred during an observation, it was coded as one single math activity rather than separate activities for each group of children.

Finally, those math activities that were very short in time (30 seconds or less) or did not include direct involvement by the teacher were cataloged as Miniature Specific Math Activities, or miniSMAs, and were coded for topic and organizational makeup. Both treatment and control classrooms were evaluated using the COEMET. The scoring for the COEMET is described in the appropriate section later in this report.

- *Near Fidelity*. The Near Fidelity (Sarama, Clements, Starkey, Klein, & Wakeley, 2008) instrument assesses the degree to which teachers in the experimental classrooms implemented components of the *Building Blocks* curriculum. It includes items in five separate settings – General Curriculum, Hands-On Centers, Whole Group, Small Group, & Computers – primarily using 5-option Likert scales. Unlike the COEMET, this fidelity instrument was

only used in the treatment classrooms. The scoring for the Near Fidelity is described in the appropriate section later in this report.

Data Analysis & Results

Descriptives. Descriptive statistics for each time point on the REMA are displayed in Table 2 separated by experimental condition.

Table 2. REMA T-Score Descriptive Statistics (Unadjusted, Unimputed) by Condition and Site, Complete Sample

	Treatment				Control			
	Min	Max	M	SD	Min	Max	M	SD
T1 (Beginning Pre-K)								
Buffalo (N=946)	10.58	52.64	38.10	5.89	10.58	51.02	38.72	5.43
Boston (N=359)	10.58	58.94	39.21	6.23	17.66	57.27	39.85	6.56
Nashville (N=431)	5.69	54.49	38.04	5.97	13.26	51.61	37.77	5.62
Total (N=1736)	5.69	58.94	38.34	6.00	10.58	57.27	38.56	5.72
T2 (End Pre-K)								
Buffalo (N=946)	31.35	61.58	47.76	4.44	26.40	55.77	44.70	5.07
Boston (N=359)	32.39	63.23	49.48	4.87	36.98	65.51	48.24	4.95
Nashville (N=409)	30.93	57.35	46.86	4.91	21.71	53.47	43.68	5.51
Total (N=1714)	30.93	63.23	48.00	4.71	21.71	65.51	44.91	5.42
T3 (End K)								
Buffalo (N=579)	37.55	68.88	53.08	4.71	39.94	63.36	52.04	4.47
Boston (N=200)	45.44	68.13	55.40	4.12	44.04	67.31	55.54	5.27
Nashville (N=412)	34.53	65.05	52.28	4.86	37.97	61.33	51.81	4.51
Total (N=1191)	34.53	68.88	53.23	4.78	37.97	67.31	52.50	4.79
T4 (End 1 st)								
Buffalo (N=530)	46.63	72.89	60.41	4.95	45.62	74.18	59.93	4.56
Boston (N=191)	52.37	74.95	62.30	4.44	49.46	75.31	62.79	5.29
Nashville (N=407)	37.85	72.35	58.33	5.17	45.27	68.45	57.68	4.74
Total (N=1128)	37.85	74.95	60.01	5.14	45.27	75.31	59.56	5.05

Missing Data. Missing data imputation techniques were used after baseline demographic differences were examined for missing data on variables other than the outcome of interest. The analytic sample included no missing data on the outcome variables and relatively little on the covariates (e.g., 2.6% at Time 2). With so little missing data, we saw no need to use multiple imputation to add variance to the missing values and thus imputed only once for each of four analytic samples (overall and each site).

Baseline Equivalence. The success of the randomization was examined by testing for significant condition differences in children’s baseline characteristics in each of the analytic samples. Multilevel regressions were run to test for the significance of any treatment-control baseline differences. Results are in Table 3 below. Additionally, standardized mean

difference effect sizes for the pretest variable in each complete sample are shown in Table 4 by site and overall. The only significant difference involved the test lag variable. It was -0.46, that is, the treatment group was assessed half a day later than the control group. It was statistically significant at an Alpha of 0.05 given such large sample size, but was not practical significance¹. All demographic baseline characteristics and pretest scores were included in all the analysis models presented in this summary, whether or not they were significantly different by treatment group, to conservatively account for even the smallest baseline differences.

Table 3. Complete and Common Samples: Baseline Covariate Comparison Across Conditions

Covariate	Group Mean Difference (T-C)	SE	p-value
Time 2 Complete Sample			
EMA Pretest	-0.24	0.57	.678
Age at Pretest	-0.67	0.52	.200
Test Lag from School Start to Pretest	-0.46*	0.06	.000
Ethnicity: Black	0.01	0.09	.950
Ethnicity: White	0.00	0.05	.979
Ethnicity: Hispanic	-0.02	0.07	.734
English Language Learner Status	-0.06	0.06	.374
Gender (Male)	0.00	0.03	.937
Parent Education	0.17	0.09	.066
Time 3 Complete Sample			
EMA Pretest	-0.41	0.68	.553
Age at Pretest	-0.60	0.56	.290
Test Lag from School Start to Pretest	-0.39*	0.07	.000
Ethnicity: Black	0.06	0.10	.588
Ethnicity: White	0.00	0.06	.949
Ethnicity: Hispanic	-0.06	0.08	.476
English Language Learner Status	-0.08	0.07	.259
Gender (Male)	0.01	0.03	.695
Parent Education	0.06	0.10	.527
Time 4 Complete Sample			
EMA Pretest	-0.21	0.68	.758
Age at Pretest	-0.54	0.55	.328
Test Lag from School Start to Pretest	-0.38*	0.07	.000
Ethnicity: Black	0.06	0.10	.578
Ethnicity: White	0.00	0.06	.975
Ethnicity: Hispanic	-0.05	0.08	.508
English Language Learner Status	-0.08	0.07	.282

¹ The test lag between treatment and control groups for Buffalo, Boston, and Nashville samples are: -0.71 (p <0.0001), -0.36 (p <0.0001), and -0.12 (p =0.13), respectively.

Covariate	Group Mean	SE	p-value
Gender (Male)	0.01	0.03	.844
Parent Education	0.07	0.10	.485
Common Sample			
EMA Pretest	-0.17	0.68	.800
Age at Pretest	-0.52	0.55	.343
Test Lag from School Start to Pretest	-0.39*	0.07	.000
Ethnicity: Black	0.06	0.10	.593
Ethnicity: White	0.00	0.06	.978
Ethnicity: Hispanic	-0.06	0.08	.495
English Language Learner Status	-0.08	0.07	.269
Gender (Male)	0.01	0.03	.836
Parent Education	0.06	0.11	.543

Note. Complete Sample N's: T2=1714, T3=1192, T4=1129; Common Sample N=1105

*p<.05

Table 4. Complete and Common Samples: Baseline REMA Pretest Randomization Check (Overall and by Site)

Site	Group Mean Difference (T-C)	SE	p-value	Effect Size
Time 2 Complete Sample				
Overall	-0.24	0.57	.678	-0.05
Buffalo	-0.68	0.75	.374	-0.13
Boston	-0.65	1.49	.671	-0.12
Nashville	0.28	0.97	.777	0.05
Time 3 Complete Sample				
Overall	-0.41	0.68	.553	-0.08
Buffalo	-0.58	0.90	.529	-0.12
Boston	-0.56	2.22	.807	-0.12
Nashville	0.04	1.03	.969	0.01
Time 4 Complete Sample				
Overall	-0.21	0.68	.758	-0.04
Buffalo	-0.36	0.92	.704	-0.07
Boston	-0.47	2.14	.832	-0.09
Nashville	0.21	1.04	.845	0.04
Common Sample				
Overall	-0.17	0.68	.800	-0.03
Buffalo	-0.36	0.92	.704	-0.06
Boston	-0.47	2.14	.832	-0.08
Nashville	0.33	1.06	.763	0.05

Note. Complete Sample N's: T2=1714, T3=1192, T4=1129; Common Sample N=1105

Research Question 1: Immediate and Longitudinal Intervention Effects. To examine the immediate and longitudinal effects of the treatment on children’s math gain, children’s outcome scores were regressed on experimental condition, controlling for their beginning of Pre-K scores, age at posttest, gender, ethnicity, English language status, test lag between school start and pretest, parent education, and interval between testing periods, in multilevel models with random intercepts and fixed slopes. Children were nested within their Pre-K classrooms and schools. There were no classroom-level covariates, but site (i.e. state) was included as a school-level covariate. T-scores were used in the REMA model.

The analysis results for Research Question 1 are shown in Table 4 below, and comprehensive tables showing the full analysis results for each of these outcomes are in the appendices (Tables A4-A6 for the complete sample and Tables B4-B6 for the common sample). These results are also illustrated in Figures A1-2 and B1-2 in those appendices. Though the parameter estimates and associated standard errors and p-values are slightly different for the common sample, the effect sizes for both samples are presented in Table 5 for comparison purposes. Note that these analyses were run with the original pretest scores at Level 1. An analysis was also run with pretest school-centered at Level 1 and the school mean pretest scores at Level 3; the results were virtually the same.

Table 5. Complete and Common Samples: Main Effects Results for Each Time Point

Outcome	B	SE	p-value	Effect Size (Complete Sample)	Effect Size (Common Sample)
T2 (End of Pre-K)	2.74*	0.34	.000	.53	.56
T3 (End of Kindergarten)	0.74*	0.27	.009	.15	.17
T4 (End of 1 st Grade)	0.51	0.31	.106	.10	.10

Note. Complete Sample N's: T2=1714, T3=1192, T4=1129; Common Sample N= 1105

* $p < .05$

Research Question 2: Variation in Effects across Sites. The same model as the one above was run again for all sites combined with the Complete Sample, but including the interaction of site and condition. The overall interaction effect was marginally significant at T2 ($F=2.65, p=.078$) but not significant at T3 ($F=0.70, p=.501$) or T4 ($F=0.25, p=.777$). The main effects for TRIAD, therefore, were not significantly different across sites, with the possible exception of the end of Pre-K. For examination of the contrasts between sites, Tennessee was used as the reference group; the results are presented in Table 6. The models were rerun with Boston as the reference group to obtain statistics for the third comparison. Full model results for the Complete Sample can be found in Tables A7-9 in Appendix A and, for the Common Sample, in Tables B7-9 in Appendix B. These results are also illustrated in Figures A3 and B3 in the appendices.

Table 6. Individual Site Contrasts from the Model Including the Site x Condition Interaction (Complete Sample)

Contrast	T2 (End of PK)			T3 (End of K)			T4 (End of 1 st)		
	b	SE	p	b	SE	p	b	SE	p
NY v. TN	0.01	0.75	.989	0.59	0.60	.334	0.05	0.69	.940
MA v. TN	-1.71 [†]	0.89	.056	-0.11	0.76	.890	-0.52	0.86	.550
NY v. MA	1.73*	0.80	.034	0.69	0.72	.340	0.57	0.83	.495

Note. Complete Sample N's: T2=1714, T3=1192, T4=1129 *p<.05, †p<.10

In addition, the main effects analysis was conducted for each site individually; the results are presented in Table 7.

Table 7. Condition Effects on REMA Scores for Sites Analyzed Separately

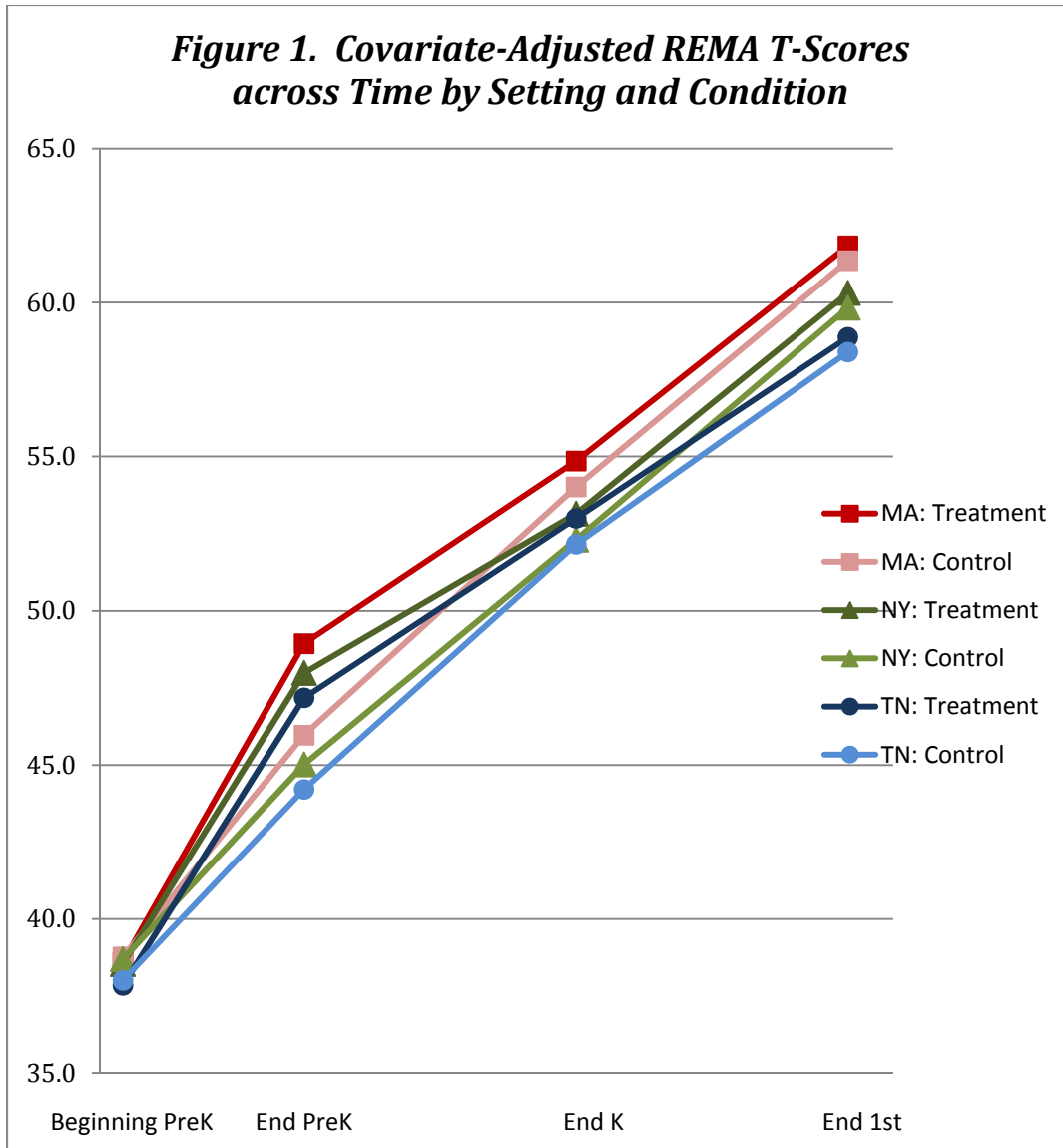
Outcome	Buffalo			Boston			Nashville		
	b (Complete Sample)	Effect Size (Complete Sample)	Effect Size (Common Sample)	b (Complete Sample)	Effect Size (Complete Sample)	Effect Size (Common Sample)	b (Complete Sample)	Effect Size (Complete Sample)	Effect Size (Common Sample)
End of PK	3.07**	.63	.62	1.44*	.29	.48	3.19**	.58	.58
End of K	0.87	.19	.22	0.91	.19	.17	0.62	.13	.15
End of 1st	0.55	.12	.12	0.55	.11	.11	0.75	.15	.15

Notes. NY Complete Sample N's: T2=946, T3=579, T4=530; Common Sample N=530; MA Complete Sample N's: T2=359, T3=200, T4=191; Common Sample N=191; TN Complete Sample N's: T2=409, T3=413, T4=408; Common Sample N=384.

*p<.05, **p<.01

Tables with the results from the full models for the analyses above are in the Appendices in Tables A10-12 for the Complete Sample and in Tables B10-12 for the Common Sample.

Figure 1 is a line graph displaying the covariate-adjusted REMA means at each time point for each condition and site for the Common Sample.



	Beginning PK	End PK	End K	End 1 st
MA: Treatment	38.6	48.9	54.9	61.8
MA: Control	38.8	46.0	54.0	61.4
NY: Treatment	38.5	48.0	53.1	60.3
NY: Control	38.7	45.0	52.3	59.8
TN: Treatment	37.8	47.2	53.0	58.9
TN: Control	38.0	44.2	52.2	58.4

Research Question 3: Curriculum Effects for Subgroups of Children. To examine whether there were differential effects of the treatment for different subgroups of children, the same model as in Research Question 1 was run but with cross-level condition x moderator interactions (in separate analyses for each moderator). Moderators of interest included gender, English Language Learner (ELL) status, pretest scores, age at pretest, and ethnicity. Results are shown in Table 8.

Table 8. Condition x Child Characteristic Interactions in the Complete Sample (each moderator examined in a separate analysis)

Moderator	End of PK Outcome		End of K Outcome		End of 1 st Outcome	
	b	p-value	b	p-value	b	p-value
Gender (Male)	0.24	.545	-0.16	.714	-0.23	.623
ELL	0.06	.914	-0.30	.649	-0.32	.661
Pretest	-0.13**	.000	-0.04	.288	0.00	.938
Age at Posttest	-0.13*	.011	-0.08	.152	0.01	.860
Black v. NonBlack	1.23**	.007	0.56	.239	0.69	.192
White v. NonWhite	-0.52	.376	0.06	.926	0.03	.962
Hispanic v. NonHispanic	-0.91	.105	-0.16	.799	-0.43	.527

* $p < .05$, ** $p < .01$

Summary of the results of the moderator analyses:

- TRIAD treatment was more effective for students who had lower skills at the beginning of Pre-K than it was for students with higher entering skills. Effects at the end of K and 1st grade were not different for children entering Pre-K with different pretest scores.
- Treatment was more effective for younger students than it was for older students. Effects at the end of K and 1st grade were not different depending on age at entry to Pre-K.
- Treatment was more effective for Black students than it was for Non-Black students. Effects at the end of K and 1st grade were not different depending on ethnicity.
- The effect of the treatment on outcomes at any year was not different for boys compared with girls, for English Language Learners compared with Native English speakers, or for White or Hispanic students compared with those who were not White or Hispanic. In analyzing ELL as a moderator, analyses were also conducted by excluding the variable for Hispanic ethnicity in the model. The interaction effect estimate was slightly different from that reported above, but still not statistically significant.

Tables including the results from the full models for the analyses of the complete samples above can be found in Tables A13-A15 in the Appendix A. The moderator analysis results for the common samples are in Tables B13-B15 in the Appendix B and are substantially similar to those found for the complete samples.

We also tested three-way interactions for condition x site x each of the significant moderators from the above analyses (pretest, age, and Black v. Non-Black) in separate analyses to determine whether the moderator effects differed across sites. The only analysis in which the three-way interaction was significant was for the pretest moderator (full results in Table A16 in Appendix A). The results indicated that the moderator effect of pretest varied significantly by site at all three time points. To explore this further, we analyzed the moderator effect of pretest within each site separately. At Time 2, the relationship between condition and Pre-K outcome differed significantly depending on pretest scores in the Nashville site ($b=-0.18$, $SE=0.07$, $p=.010$) and marginally for the Buffalo site ($b=-0.09$, $SE=0.05$, $p=.053$), but not for the Boston site ($b=0.00$, $SE=0.08$, $p=.999$). The direction of the effect for both the Nashville and Buffalo sites was the same as in the combined analysis—children with lower pretest scores showed larger gains during the Pre-K year. Full model results for each time point can be found in Tables A17-19 in Appendix A.

Research Question 4: Effects of Math Environment and Fidelity of Implementation on Gain. The two measures of interest were the COEMET, used in both treatment and control classrooms, and the Near Fidelity measure, used only in the treatment classrooms.

COEMET Exploration

Table 9 below shows descriptive information for the original COEMET variables.

Table 9. Descriptive Statistics for Pre-K Classrooms on the COEMET by Condition and Site

	Treatment				Control			
	Min	Max	M	SD	Min	Max	M	SD
Classroom Culture								
Buffalo	2.50	4.61	4.07	0.37	2.19	4.61	3.57	0.60
Boston	3.50	4.44	4.01	0.26	2.67	4.06	3.32	0.43
Nashville	2.90	4.59	4.07	0.41	2.49	4.04	3.28	0.46
Total	2.50	4.61	4.06	0.35	2.19	4.61	3.42	0.54
Number SMA's								
Buffalo	1.00	6.50	3.45	1.11	0.50	3.50	2.13	0.85
Boston	2.00	7.50	4.83	1.35	1.50	6.00	3.70	1.49
Nashville	1.00	4.00	2.94	0.79	0.33	4.33	1.73	0.88
Total	1.00	7.50	3.69	1.30	0.33	6.00	2.30	1.23

Quality SMA's								
Buffalo	2.90	4.61	3.90	0.23	3.24	4.18	3.76	0.22
Boston	3.35	4.47	3.65	0.24	2.43	3.78	3.21	0.37
Nashville	3.27	4.09	3.63	0.23	2.15	3.67	3.23	0.35
Total	2.90	4.61	3.79	0.27	2.15	4.18	3.48	0.40
Number Mini SMA's								
Buffalo	0.50	7.00	3.00	1.46	0.50	4.50	2.27	1.01
Boston	1.00	6.50	3.52	1.50	0.50	3.50	2.20	0.89
Nashville	2.33	15.67	8.96	3.83	2.00	11.67	5.47	2.63
Total	0.50	15.67	4.21	3.07	0.50	11.67	3.32	2.28

Notes. Buffalo Control N=24, Buffalo Treatment N=51; Boston Control N=10, Boston Treatment N=21; Nashville Control N=17, Nashville Treatment N=16

The descriptive statistics for the COEMET measures showed some notable site differences. To explore these differences further, we conducted a two-level HLM by including Site, Condition, and the interaction term of Site by Condition to predict the COEMET measures. There were significant site differences for the Number SMAs, Quality of SMAs, and Number of Mini SMAs. The interaction terms for Condition x Site for SMA quality and Number of Mini SMA's were also statistically significant (the analysis results are in Tables A20 and A21 in Appendix A). The means for each site on each of the COEMET measures are shown above in Table 8. Though actual differences as large as some of those shown between the sites on the classroom events the COEMET represents are possible, the differences could also result from variations among observers at the different sites in the ways they made and interpreted their COEMET observations. Unable to disentangle those, we opted to site-center the COEMET variables in our analyses so that they varied within site but not between sites.

A composite COEMET variable was then created from the components of the measure through principal factor analysis using classroom-level scores. The four COEMET measures were first rescaled using site-mean centering, i.e., subtracting the site means from the original scores for each classroom in each site. The Cronbach alpha for the four subscales was 0.75, and a principal factor analysis showed that a one factor-solution provided an adequate representation of them (the factor loadings are in Table A22 in Appendix A). The factor score from that analysis was used as the composite COEMET variable for summary analysis.

This composite COEMET score was first analyzed as the dependent variable in a two-level HLM with Condition, Site, and the Condition x Site interaction as predictors (shown in the last columns of Tables A20 and A21). The results showed a significant difference for condition, with classrooms in the treatment condition showing higher scores than the control condition. The Condition x Site interaction was not statistically significant,

indicating no reliable difference between sites in the strength of the relationship between Condition and the COEMET composite variable.

To examine the effect of general mathematics instruction on children's math performance, the same analysis used to investigate Research Question 1 was run again, but with the COEMET composite as the key predictor in place of experimental condition. The COEMET composite was significantly and positively related to children's REMA gains during Pre-K with child-level demographic characteristics controlled ($F = 35.11, p < .0001$) (Table A23 in Appendix A). Moreover, the Site x COEMET interaction was not significant, indicating that the influence of the COEMET was similar across sites. Using the four site-centered subscales of the COEMET in place of the composite in that analysis showed that strongest independent relationship to REMA gain was made by the Quality of SMA subscale, the only one of the four to show a statistically significant individual relationship. The weakest independent relationship was shown by the number of SMAs (Table A24 in Appendix A).

The Pre-K COEMET composite also significantly and positively predicted children's REMA gains through Kindergarten ($F=5.51, p=.013$; Table A25 in Appendix A). Once again, the strongest independent contribution to the effect came from the SMA quality rating (Table A26 in Appendix A). Though positively related, Pre-K COEMET scores were not significantly predictive of children's gain through first grade (Tables A27 and A28 in Appendix A).

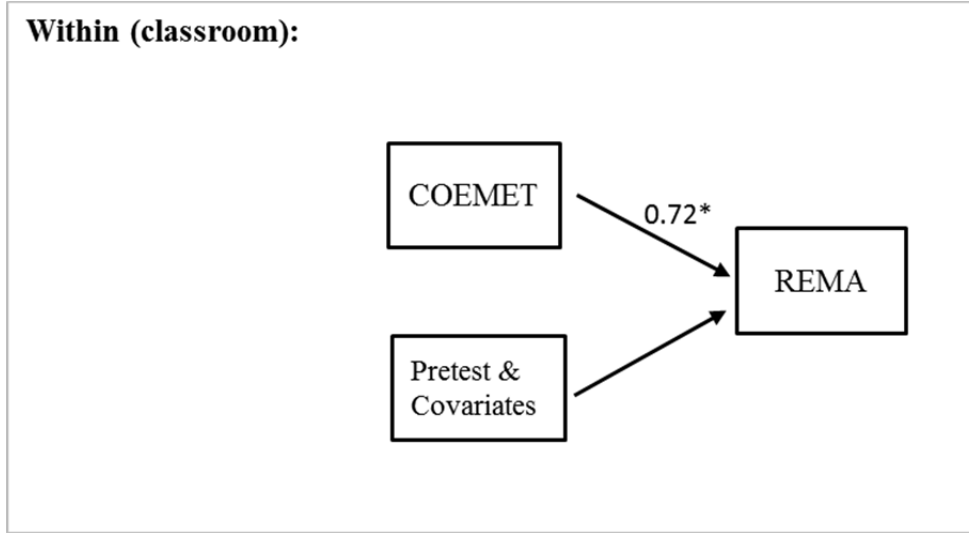
An additional analysis was done to examine the extent to which the Pre-K COEMET composite score moderated the treatment effect on REMA gain through Pre-K (Time 2). That is, this analysis asked whether the treatment effect varied depending on the overall COEMET score. In a three-level HLM including Condition, COEMET Composite, and the Condition x COEMET interaction, the interaction term was not statistically significant (Table A29 in Appendix A), indicating that the effects of the intervention on Pre-K gain did not depend on COEMET Composite scores.

The final analysis with the Pre-K COEMET composite variable was to examine its role as a mediator of the treatment effect. This was done with a multilevel SEM model implemented with MPlus. Student-level data were aggregated to the classroom level, resulting in 139 classrooms nested within 59 schools. A two-level SEM was then fit with the lower (within) level representing classrooms within schools and the higher level (between) representing between school relationships. Note that the predictor of interest (Treatment) is at the higher level, which is the one where mediation is hypothesized to occur.

The between-level model revealed a significant indirect effect (0.98, $p=0.01$, 95% confidence interval: (0.23, 1.72)). This is the mediational pathway indicating that school-level treatment increased school-level REMA by, to at least some extent, changing school-

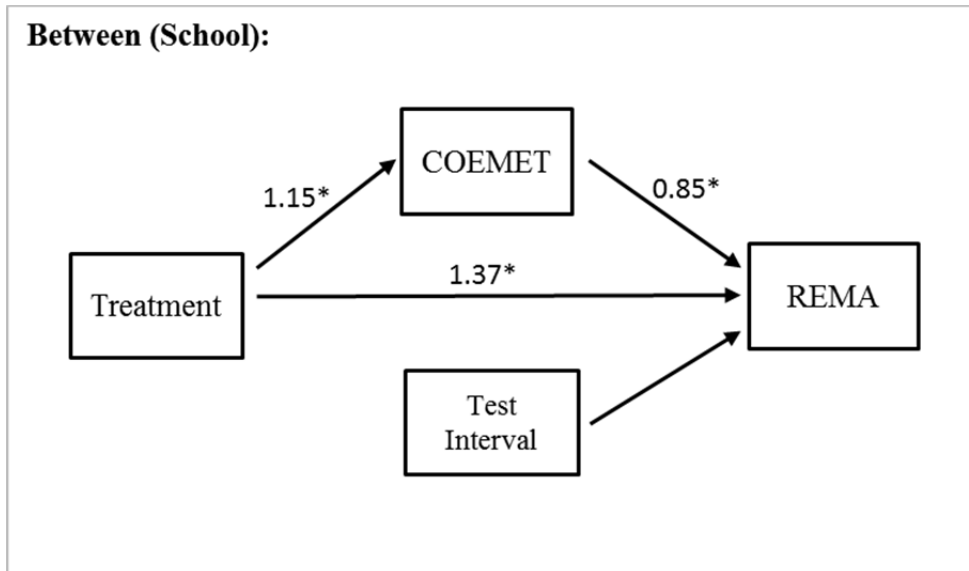
level COEMET. The results of the best fitting model are summarized in the charts below (Figures 2 and 3).

Figure 2. Level 1 of SEM Mediation Analysis: Classrooms within Schools



* $p < .05$

Figure 3. Level 2 of SEM Mediation Analysis: Between Schools



* $p < .05$

The detailed results for this analysis are provided in Tables A30-32 in Appendix A. An alternate analysis of the COEMET mediation effect was also conducted using the Barron and Kenny framework implemented with a multilevel model in SAS, The indirect (mediation) effect was tested using the Sobel test, which indicated partial mediation by the

COEMET (indirect effect estimate = 1.19, $p < 0.0001$, 95% confidence interval: (0.18, 2.21)). This estimate of the indirect effect is somewhat larger than that found in the MPlus SEM analysis, but still substantially similar and also statistically significant.

Near Fidelity Exploration

The Near Fidelity measure included items in five subscales representing fidelity to the curriculum in different settings: General Curriculum, Hands on Centers, Whole Group, Small Group, and Computers and was only collected in TRIAD treatment classrooms. Each of these subscales contains several items; items were not constructed consistently. Some were score 0 or 1; some were scored on a 1-5 Likert scale basis. Table 10 below shows descriptive information for the original Near Fidelity variables. The scoring option that we chose for this display of descriptives involved averaging the items within each of the five settings after recoding the dichotomous items as a 2 or 4 scaling rather than a 0 or 1. To deal with missing data, we imputed missing total scores for a given classroom and setting and time point to represent the lowest possible average score. Missing data at the item-level were ignored.

As with the COEMET, we did not feel confident that differences in these measures across sites represented only true site differences that were not confounded with observer differences across sites. We therefore used site-centered values in all the analyses; that is, the measures were scaled so that they represented variation within sites but not between sites.

Table 10. Descriptive Statistics for Treatment Classrooms on the Near Fidelity Subscales by Site

	Min	Max	M	SD
General Curriculum				
Boston	2.80	4.60	4.00	0.38
Buffalo	2.47	4.40	3.78	0.41
Nashville	3.13	4.47	3.86	0.35
Total	2.47	4.60	3.85	0.40
Hands On Centers				
Boston	1.30	4.25	3.51	0.76
Buffalo	1.53	4.17	3.47	0.61
Nashville	1.81	3.83	3.12	0.60
Total	1.30	4.25	3.41	0.65
Whole Group				
Boston	3.36	4.64	4.02	0.31
Buffalo	2.24	4.57	3.87	0.45
Nashville	2.95	4.48	3.86	0.42

Total	2.24	4.64	3.90	0.41
Small Group				
Boston	3.39	4.43	3.99	0.26
Buffalo	1.10	4.29	3.61	0.53
Nashville	1.65	4.27	3.10	0.81
Total	1.10	4.43	3.61	0.61
Computers				
Boston	3.02	4.29	3.82	0.39
Buffalo	1.97	4.13	3.49	0.53
Nashville	2.79	4.18	3.54	0.41
Total	1.97	4.29	3.58	0.49

Notes. Total N=88; Buffalo N=51; Boston N=21; Nashville N=16

Several scoring options were explored that dealt with missing data and scoring of items in different ways. Options included:

- Option 1:
 - Scoring Choice: average of relevant 2- or 5-point scale items within each setting
 - Recoding Choice: None (0-1 items left as 0,1)
 - Missing Data Choice: missing data ignored
- Option 2:
 - Scoring Choice: unrotated factor scores from 1st factor of each setting
 - Principal components factor analyses within each subscale (uncentered) showed reasonable coherence at the subscale level and produced factor scores for each of the subscales. These factor scores, as shown in Table A33 in Appendix A, were then site-centered for use in analyses at the subscale level. Although 3 settings loaded on more than one factor, we chose the first factor loadings since most items generally loaded most highly on that factor.
 - Recoding Choice: None (0-1 items left as 0-1)
 - Missing Data: missing data ignored
- Option 3:
 - Scoring Choice: average of relevant 2- or 5-point scale items within each setting
 - Recoding Choice: 0-1 items recoded as 2,4
 - Missing Data Choice: imputed missing total scores for a given classroom and setting and time point to represent lowest possible average score; missing item-level data ignored

Zero-order correlations between mean classroom residualized Pre-K REMA gain and Near Fidelity setting scores using the different scoring options can be found in Table A33 in

Appendix A. Neither scoring method (average v. factor scores) nor recoding (0,1 or 2,4 for dichotomous items) made much of a difference in correlations with gain. How missing data was treated made a slight difference in correlations with gain. By a small margin, the Option 2 scoring showed the largest correlation with gain, and we selected it for use in further analyses.

To provide an overall summary score for Near Fidelity, a composite score was created from these the setting factor scores. After exploration of several options, a simple *Near Fidelity Additive Score* was selected as the version that best represented differences in fidelity between classrooms. This Near Fidelity Additive Score was created by dichotomizing the site-centered factor scores for each subscale at the median into low and high values scored simply as 0 for low and 1 for high. These binary subscale scores were then summed across the five subscales to produce a composite overall score that ranged from 0 to 5.

Table 11 below presents descriptive information for this Near Fidelity Additive composite scale overall and by site.

Table 11. Near Fidelity Additive Score Descriptives by Site and Overall

	N (classrooms)	Min	Max	M	SD
Buffalo	51	0.00	5.00	2.65	1.60
Boston	21	0.00	5.00	2.29	1.95
Nashville	16	1.00	5.00	2.50	1.41
Total	88	0.00	5.00	2.53	1.65

To examine the effect of fidelity to the Building Blocks curriculum in the TRIAD project on REMA gain, the same analysis model used to investigate Research Question 1 was run again, but with the Near Fidelity Additive Score entered in place of condition, and the interaction of site and the Near Fidelity score (with a reduced sample, as the Near Fidelity measure was used only in treatment classrooms). Child-level data included all children in the treatment classrooms (both treatment conditions). The Near Fidelity Additive Score was significant as a predictor of Pre-K REMA gains ($b=0.22$, $SE=0.10$, $p=.028$; full results are in Table A35 in Appendix A). Because of missing data on some of the Near Fidelity measures, we also ran this analysis with only those classrooms that had complete data across all 5 settings. The results were similar in pattern.

To consider whether any of the subscales for the different settings had individual relationships with REMA gain, we also used the factor score for each setting to predict Pre-K gain. Because of the high correlations across settings for those factor scores, identical models were run predicting Pre-K gain from each Near Fidelity subscale separately. The results are shown in Table 12 below, and full model results are in Table A36 in Appendix A.

Table 12. Predicting Pre-K gain from Separate Near Fidelity Setting Factor Scores

Setting	b	SE	p-value
General Curriculum	0.50*	0.18	.008
Hands On Centers	-0.03	0.18	.861
Whole Group	0.39†	0.20	.050
Small Group	0.34†	0.20	.097
Computers	0.37	0.25	.139

* $p < .05$, † $p < .10$

As seen in Table 12, the site-centered subscale factor scores for the General Curriculum, Whole Group settings, and Small Group settings were positively and at least marginally significantly related to REMA gains when analyzed as individual predictors.

The same analyses were rerun only on children in classrooms with non-missing scores for each setting (N=992). The results, presented in Table A37 in Appendix A, were slightly different with this reduced sample; only General Curriculum and Whole Group scores were marginally related to child gain.

Exploration of the COEMET and Near Fidelity Measures in Combination

After examining the COEMET and Near Fidelity classroom measures of math activities separately, we examined the correlations between them overall and by site across the treatment classrooms. Those correlations are shown in Table 13 below and were statistically significant overall and at the Buffalo and Nashville sites individually. However, the magnitude of these correlations was not so great that the measures are empirically redundant. There is thus latitude for them to have independent effects on REMA gains across the classrooms.

Table 13. Correlations between the COEMET Composite and Near Fidelity Composite across Treatment Classrooms (Overall and by Site)

Site	N	r	p-value
All Classrooms			
Overall	88	.313*	.003
Buffalo	51	.318*	.023
Boston	21	.250	.274
Nashville	16	.533*	.034
Classrooms with Complete Data for All Settings			
Overall	77	.309*	.006
Buffalo	45	.297*	.048
Boston	17	.345	.175

Nashville	15	.482 [†]	.069
-----------	----	-------------------	------

* $p < .05$, [†] $p < .10$

We then addressed the question of whether specific adherence to the curriculum in the treatment classrooms (Near Fidelity) was independently related to REMA gains above and beyond the influence of the general nature of math activities in those classrooms (COEMET). To examine these relationships, we entered the COEMET composite and the Near Fidelity composite in an analysis predicting Pre-K REMA gain. With the influence of the COEMET controlled, there was no significant relationship between the Near Fidelity Additive composite and the REMA outcomes ($b=0.11$, $SE=0.09$, $p=.246$). The same result appeared when we limited the analysis to those classrooms that had complete Near Fidelity data for all 5 settings ($b=0.08$, $SE=0.10$, $p=.439$). Full results of these two models are displayed in Table A38 in Appendix A.

Summary of Results by Research Question

1. *What are the immediate and long-term effects of the intervention on children’s math skills?*

There were strong immediate positive effects of the intervention on children’s REMA gains during the Pre-K year. Children who were exposed to the TRIAD model in Pre-K made greater gains on the REMA than similar children who did not have such exposure. In Kindergarten, differences were considerably smaller but still significant, favoring the treatment group. In first grade there were no significant differences in REMA gain between children in the treatment and control conditions.

2. *How much variation was there in effects across sites?*

Intervention effects were fairly similar across sites, with some difference at the end of Pre-K. Generally, in Pre-K, Buffalo had the strongest intervention effects, with Nashville a close second and Boston a third². However, differences between these site effect sizes were only marginally significant ($p < .10$) at the end of Pre-K. In Kindergarten and the first grade, the effects are all small and not statistically different across three sites. The overall intervention effect for the sample in Kindergarten found in the exploration of Research Question 1 appeared to have been driven by Buffalo, the only site to have significant intervention effects on assessments at the end of Kindergarten.

3. *Were curricular effects different for different subgroups of children?*

² The effect sizes at the end of Pre-K were different between the complete sample and common sample for Boston. Further exploration indicated that the analytic samples varied a lot between two samples, e.g., the retention rate (the proportion of the common sample in the complete sample) for the treatment group in Boston was 78.5% where the retention rates were 86.9% and 94.3% for the Buffalo and Nashville samples. The big variation in two analytic samples in Boston resulted in different point estimates of the impact and the pooled standard deviations.

Several child-level moderators were tested to evaluate whether the effects of the intervention were differentially related to child characteristics. Curricular effects did not differ significantly depending on gender, ELL status, White versus Non-White, or Hispanic compared to Non-Hispanic. However, differences were found regarding pretest scores, age, and Black versus Non-Black. The intervention was more effective for children with low pretest scores compared to those with higher pretests, children who entered Pre-K at a younger age compared to those who were older at the start of school, and Black children compared to Non-Black children. These moderator effects were generally true across all the sites with the exception of pretest; in Boston, there were no significantly different curricular effects depending on a child's entering scores.

4. *What are the effects of the math environment and the fidelity of implementation on children's immediate and long term math gains, overall and across sites?*

We found positive relationships between classroom-level variables related to how much mathematics was observed and children's gain. Higher math environment (COEMET) composite scores, measured in both treatment and control classrooms, were correlated with greater REMA gain during Pre-K and Kindergarten in all three sites. While the actual number of math activities had the weakest relationship with classroom REMA gain, the quality of activities observed had the strongest relationship to gains.

In addition to the main effect of math environment, we examined the interplay between math environment quality and treatment. Although COEMET scores predicted greater REMA gain, they were not found to *moderate* the effects of the curriculum. The effect of the intervention on REMA gain was not different depending on classroom COEMET scores. However, COEMET scores were a significant *mediator* of mean REMA gain. Mediator analyses showed that the intervention was related to greater REMA gain at least partially through its effect on math environment quality scores.

We also examined fidelity of implementation in treatment classrooms with the Near Fidelity instrument. Near Fidelity composite scores significantly predicted greater REMA gain in Pre-K. By looking at the individual settings of the instrument, we found that scores in General Curriculum, Small Groups, and Computer settings were driving the overall positive prediction. It should be noted, however, that missing data made some slight differences. There were several missing data issues with the Near Fidelity measure, both at the item level and the setting level. When we considered only those classrooms that had valid data for all settings, the overall effect of the Near Fidelity composite was the same, but the look by setting was different. When we tested the prediction of Near Fidelity scores to Pre-K REMA gain controlling for COEMET scores, implementation fidelity no longer significantly predicted gain.

References

- Clements, D. & Sarama, J. (June 2006). *Scaling Up TRIAD: Teaching Early Mathematics for Understanding with Trajectories and Technology*, proposal funded by the Institute for Education Sciences.
- Clements, D., & Sarama, J. (2011). Early childhood mathematics intervention. *Science*, 333, 968-970.S
- Clements, D., Sarama, J., & Liu, X. (July 2008). Development of a measure of early mathematics achievement using the Rasch model: the Research-Based Early Math Assessment. *Educational Psychology*, 28 (4), 457-482.
- Sarama, J., & Clements, D. H. (2007). Manual for classroom observation (COEMET) – Version 3. Unpublished version.
- Sarama, J., Clements, D. H., Starkey, P., Klein, A., & Wakeley, A. (2008). Scaling up the implementation of a pre-kindergarten mathematics curriculum: Teaching for understanding with trajectories and technologies. *Journal of Research on Educational Effectiveness*, 1, 89-119.

Appendices

Appendix A

Table A1: Sample Sizes by Time (Series 1 – complete sample)

Time 2

Site	N1 (student)	N2 (class)	N3 (school)
Buffalo	946	75	25
Boston	359	31	18
Nashville	409	33	16
Total	1714	139	59

Time 3

Site	N1 (student)	N2 (class)	N3 (school)
Buffalo	579	49	18
Boston	200	21	12
Nashville	413	33	16
Total	1192	103	46

Time 4

Site	N1 (student)	N2 (class)	N3 (school)
Buffalo	530	49	18
Boston	191	21	12
Nashville	408	33	16
Total	1129	103	46

Table A2: Descriptive Statistics by Condition: Child Level Variables (Series 1)

Variable	Control		Treatment group	
	Mean	SD	Mean	SD
<i>Time 2</i>				
T-Score - Rasch Converted Pretest (T1)	38.53	5.74	38.36	6.01
T-Score - Rasch Converted Posttest (T2)	44.91	5.42	48.00	4.71
Race: Black	0.57	0.50	0.60	0.49
Race: White	0.15	0.36	0.17	0.37
Race: Hispanic	0.22	0.41	0.17	0.38
PK ELL Status	0.20	0.40	0.13	0.33
Male	0.48	0.50	0.48	0.50
Parent highest educational level	1.35	0.96	1.55	0.91
Child Age at REMA T2 (End of PK)	60.11	4.07	59.87	4.10
Age interval (month) from REMA T2	-7.15	0.66	-7.81	0.57
Test Lag between school start and pretest in month	1.24	0.47	0.76	0.50
<i>N</i>	576		1138	
<i>Time 3</i>				
T-Score - Rasch Converted Kindergarten (T3)	52.48	4.79	53.23	4.78
Child Age at REMA T3 (End of K)	72.10	4.09	71.80	4.09
Age interval (month) from REMA T3	-19.13	0.59	-19.55	0.60
<i>N</i>	561		631	
<i>Time 4</i>				
T-Score - Rasch Converted First Grade (T4)	59.55	5.05	60.01	5.14
Child Age at REMA T4 (End of 1st Grade)	83.90	4.01	83.74	4.03
Age interval (month) from REMA T4	-30.84	0.79	-31.38	0.88
<i>N</i>	533		596	

Table A3: Descriptive Statistics by Site and Condition: Child Level Variables (Series 1)

Variable	Buffalo				Boston				Nashville			
	Control		Treatment		Control		Treatment		Control		Treatment	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>Time 2</i>												
T-Score - Rasch Converted Pretest (T1)	38.72	5.43	38.10	5.89	39.85	6.56	39.21	6.23	37.64	5.64	38.13	6.01
T-Score - Rasch Converted Posttest (T2)	44.70	5.07	47.76	4.44	48.24	4.95	49.48	4.87	43.68	5.51	46.86	4.91
Race: Black	0.55	0.50	0.66	0.47	0.30	0.46	0.30	0.46	0.72	0.45	0.81	0.40
Race: White	0.19	0.39	0.22	0.42	0.12	0.33	0.11	0.32	0.12	0.32	0.06	0.23
Race: Hispanic	0.19	0.39	0.08	0.27	0.49	0.50	0.48	0.50	0.13	0.34	0.08	0.27
PK ELL Status	0.16	0.37	0.03	0.16	0.45	0.50	0.42	0.49	0.14	0.34	0.09	0.28
Male	0.49	0.50	0.50	0.50	0.51	0.50	0.46	0.50	0.44	0.50	0.46	0.50
Parent highest educational level	1.40	0.96	1.56	0.91	1.55	1.01	1.58	0.92	1.17	0.91	1.48	0.91
Child Age at REMA T2 (End of PK)	58.77	3.79	58.60	3.58	63.01	3.91	62.61	4.00	60.71	3.68	60.36	3.96
Age interval (month) from REMA T2	-7.09	0.61	-7.99	0.50	-7.16	0.98	-7.71	0.55	-7.22	0.54	-7.34	0.48
Test Lag between school start and pretest in month	1.38	0.49	0.67	0.50	1.12	0.33	0.78	0.44	1.08	0.44	1.03	0.47
<i>N</i>	286		660		92		267		198		211	
<i>Time 3</i>												
T-Score - Rasch Converted Kindergarten (T3)	52.04	4.47	53.08	4.71	55.54	5.27	55.40	4.12	51.78	4.52	52.28	4.86
Child Age at REMA T3 (End of K)	70.63	3.79	70.34	3.64	74.62	3.72	74.87	3.81	73.03	3.84	72.25	3.82
Age interval (month) from REMA T3	-19.00	0.51	-19.77	0.50	-18.93	0.57	-19.46	0.61	-19.38	0.63	-19.30	0.62
<i>N</i>	275		304		86		114		200		213	
<i>Time 4</i>												
T-Score - Rasch Converted First Grade (T4)	59.93	4.56	60.41	4.95	62.79	5.29	62.30	4.44	57.67	4.73	58.33	5.17
Child Age at REMA T4 (End of 1st Grade)	82.88	3.86	82.65	3.71	86.42	3.61	86.69	3.82	84.12	3.85	83.70	3.80
Age interval (month) from REMA T4	-31.17	0.61	-31.96	0.55	-30.80	0.57	-31.26	0.56	-30.43	0.89	-30.68	0.83
<i>N</i>	251		279		85		106		197		211	

Table A4: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample: Time 2 (Series 1)

Fixed Effect

Effect	b	SE	p
Intercept	21.27	3.59	0.0000
Condition	2.74	0.34	0.0000
Site 2: MA	1.60	0.48	0.0013
Site 1: NY	0.75	0.39	0.0564
Site 3: TN (reference)	0.00		
REMA pretest school-mean centered	0.42	0.02	0.0000
REMA pretest school mean	0.35	0.08	0.0000
Age at Time 2	0.16	0.03	0.0000
Age interval between Time 2 and Time 1	-0.06	0.19	0.7451
Test Lag from School Start to Pretest	-0.17	0.22	0.4453
Ethnicity: Black	-1.18	0.44	0.0080
Ethnicity: White	0.59	0.49	0.2298
Ethnicity: Hispanic	-0.49	0.47	0.2969
English Language Learner Status	0.66	0.35	0.0617
Gender (Male)	-0.51	0.18	0.0061
Parent Education	0.49	0.11	0.0000

Random Effect

Level	Variance	SE	p	ICC
Level 2 (Class)	0.58	0.27	0.0162	0.04
Level 3 (School)	0.56	0.27	0.0183	0.04
Level 1 (Student)	13.85	0.50	0.0000	NA

Random Effect - Unconditional Model

Level	Variance	SE	p	ICC
Level 2 (Class)	0.64	0.40	0.0550	0.02
Level 3 (School)	4.73	1.13	0.0000	0.18
Level 1 (Student)	21.64	0.77	0.0000	NA

Table A5: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample: Time 3 (Series 1)

Fixed Effect

Effect	b	SE	p
Intercept	30.74	5.38	0.0000
Condition	0.74	0.27	0.0093
Site 2: MA	1.82	0.41	0.0000
Site 1: NY	0.10	0.29	0.7262
Site 3: TN (reference)	0.00		
REMA pretest school-mean centered	0.39	0.02	0.0000
REMA pretest school mean	0.40	0.06	0.0000
Age at Time 3	0.11	0.03	0.0001
Age interval between Time 3 and Time 1	0.03	0.22	0.8789
Test Lag from School Start to Pretest	-0.31	0.27	0.2409
Ethnicity: Black	-2.69	0.54	0.0000
Ethnicity: White	-0.45	0.58	0.4409
Ethnicity: Hispanic	-1.74	0.57	0.0024
English Language Learner Status	0.76	0.42	0.0708
Gender (Male)	-0.23	0.22	0.2924
Parent Education	0.45	0.13	0.0003

Random Effect

Level	Variance	SE	p	ICC
Level 2 (Class)	0.36	0.27	0.0918	0.03
Level 3 (School)	0.02	0.17	0.4403	0.00
Level 1 (Student)	13.21	0.57	0.0000	NA

Random Effect - Unconditional Model

Level	Variance	SE	p	ICC
Level 2 (Class)	0.87	0.49	0.0380	0.04
Level 3 (School)	2.71	0.91	0.0014	0.12
Level 1 (Student)	19.77	0.85	0.0000	NA

Table A6: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample: Time 4 (Series 1)

Fixed Effect

Effect	b	SE	p
Intercept	44.78	6.81	0.0000
Condition	0.51	0.31	0.1062
Site 2: MA	3.02	0.47	0.0000
Site 1: NY	1.56	0.38	0.0001
Site 3: TN (reference)	0.00		
REMA pretest school-mean centered	0.40	0.02	0.0000
REMA pretest school mean	0.42	0.07	0.0000
Age at Time 2	0.04	0.03	0.1822
Age interval between Time 2 and Time 1	0.16	0.19	0.4037
Test Lag from School Start to Pretest	-0.11	0.28	0.7015
Ethnicity: Black	-2.83	0.60	0.0000
Ethnicity: White	-0.49	0.65	0.4519
Ethnicity: Hispanic	-1.52	0.64	0.0170
English Language Learner Status	0.58	0.47	0.2155
Gender (Male)	0.21	0.24	0.3859
Parent Education	0.51	0.13	0.0002

Random Effect

Level	Variance	SE	p	ICC
Level 2 (Class)	0.44	0.35	0.1055	0.03
Level 3 (School)	0.08	0.24	0.3700	0.01
Level 1 (Student)	15.14	0.67	0.0000	NA

Random Effect - Unconditional Model

Level	Variance	SE	p	ICC
Level 2 (Class)	0.77	0.55	0.0804	0.03
Level 3 (School)	4.05	1.25	0.0006	0.15
Level 1 (Student)	21.68	0.96	0.0000	NA

Table A7: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample with Condition x Site Interaction: Time 2 (Series 1)

Fixed Effect			
Effect	b	SE	p
Intercept	21.14	3.54	0.0000
Condition	3.13	0.59	0.0000
Site 2: MA	2.72	0.72	0.0003
Site 1: NY	0.69	0.56	0.2230
Site 3: TN (reference)	0.00		
Condition*site 2	-1.71	0.89	0.0565
Condition*site 1	0.01	0.75	0.9886
Condition*site 3	0.00		
REMA pretest school-mean centered	0.42	0.02	0.0000
REMA pretest school mean	0.35	0.08	0.0000
Age at Time 2	0.16	0.03	0.0000
Age interval between Time 2 and Time 1	-0.06	0.19	0.7332
Test Lag from School Start to Pretest	-0.15	0.22	0.4876
Ethnicity: Black	-1.19	0.44	0.0073
Ethnicity: White	0.58	0.49	0.2316
Ethnicity: Hispanic	-0.48	0.47	0.3037
English Language Learner Status	0.67	0.35	0.0559
Gender (Male)	-0.51	0.18	0.0055
Parent Education	0.49	0.11	0.0000

Random Effect				
Level	Variance	SE	p	ICC
Level 2 (Class)	0.57	0.27	0.0168	0.04
Level 3 (School)	0.50	0.25	0.0243	0.03
Level 1 (Student)	13.84	0.50	0.0000	NA

Type 3 Test		
Effect	F Value	p
Condition	55.12	0.0000
Site	7.26	0.0013
Condition*site	2.65	0.0783
REMA pretest school-mean centered	545.75	0.0000
REMA pretest school mean	21.37	0.0000
Age at Time 2	37.24	0.0000
Age interval between Time 2 and Time 1	0.12	0.7332
Test Lag from School Start to Pretest	0.48	0.4876
Ethnicity: Black	7.21	0.0073
Ethnicity: White	1.43	0.2316
Ethnicity: Hispanic	1.06	0.3037
English Language Learner Status	3.66	0.0559
Gender (Male)	7.72	0.0055
Parent Education	21.09	0.0000

Table A8: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample with Condition x Site Interaction: Time 3 (Series 1)

Fixed Effect			
Effect	b	SE	p
Intercept	31.29	5.45	0.0000
Condition	0.52	0.43	0.2405
Site 2: MA	1.88	0.59	0.0020
Site 1: NY	-0.19	0.42	0.6567
Site 3: TN (reference)	0.00		
Condition*site 2	-0.11	0.76	0.8897
Condition*site 1	0.59	0.60	0.3339
Condition*site 3	0.00		
REMA pretest school-mean centered	0.39	0.02	0.0000
REMA pretest school mean	0.41	0.06	0.0000
Age at Time 3	0.11	0.03	0.0001
Age interval between Time 3 and Time 1	0.06	0.23	0.7812
Test Lag from School Start to Pretest	-0.26	0.27	0.3406
Ethnicity: Black	-2.70	0.54	0.0000
Ethnicity: White	-0.49	0.59	0.4004
Ethnicity: Hispanic	-1.75	0.57	0.0023
English Language Learner Status	0.77	0.42	0.0659
Gender (Male)	-0.23	0.22	0.2943
Parent Education	0.45	0.13	0.0004

Random Effect				
Level	Variance	SE	p	ICC
Level 2 (Class)	0.36	0.27	0.0871	0.03
Level 3 (School)	0.03	0.17	0.4274	0.00
Level 1 (Student)	13.20	0.57	0.0000	NA

Type 3 Test		
Effect	F Value	p
Condition	5.27	0.0258
Site	7.19	0.0016
Condition*site	0.70	0.5009
REMA pretest school-mean centered	362.34	0.0000
REMA pretest school mean	42.20	0.0000
Age at Time 3	14.74	0.0001
Age interval between Time 3 and Time 1	0.08	0.7812
Test Lag from School Start to Pretest	0.91	0.3406
Ethnicity: Black	24.88	0.0000
Ethnicity: White	0.71	0.4004
Ethnicity: Hispanic	9.34	0.0023
English Language Learner Status	3.39	0.0659
Gender (Male)	1.10	0.2943
Parent Education	12.80	0.0004

Table A9: Fixed and Random Effects Analysis for REMA Outcome -Complete Sample with Condition x Site Interaction: Time 4 (Series 1)

Fixed Effect			
Effect	b	SE	p
Intercept	44.91	6.85	0.0000
Condition	0.59	0.50	0.2474
Site 2: MA	3.29	0.66	0.0000
Site 1: NY	1.54	0.52	0.0045
Site 3: TN (reference)	0.00		
Condition *site 2	-0.52	0.86	0.5504
Condition *site 1	0.05	0.69	0.9399
Condition *site 3	0.00		
REMA pretest school-mean centered	0.40	0.02	0.0000
REMA pretest school mean	0.42	0.07	0.0000
Age at Time 4	0.04	0.03	0.1772
Age interval between Time 4 and Time 1	0.17	0.19	0.3897
Test Lag from School Start to Pretest	-0.09	0.29	0.7438
Ethnicity: Black	-2.82	0.60	0.0000
Ethnicity: White	-0.47	0.65	0.4673
Ethnicity: Hispanic	-1.52	0.64	0.0175
English Language Learner Status	0.58	0.47	0.2142
Gender (Male)	0.21	0.24	0.3851
Parent Education	0.50	0.14	0.0002

Random Effect				
Level	Variance	SE	p	ICC
Level 2 (Class)	0.44	0.35	0.1071	0.03
Level 3 (School)	0.13	0.26	0.3025	0.01
Level 1 (Student)	15.14	0.67	0.0000	NA

Type 3 Test		
Effect	F Value	p
Condition	1.63	0.2078
site	12.80	0.0000
Condition *site	0.25	0.7765
REMA pretest school-mean centered	302.33	0.0000
REMA pretest school mean	33.80	0.0000
Age at Time 4	1.82	0.1772
Age interval between Time 4 and Time 1	0.74	0.3897
Test Lag from School Start to Pretest	0.11	0.7438
Ethnicity: Black	22.09	0.0000
Ethnicity: White	0.53	0.4673
Ethnicity: Hispanic	5.67	0.0175
English Language Learner Status	1.54	0.2142
Gender (Male)	0.76	0.3851
Parent Education	13.77	0.0002

Table A10: Fixed and Random Effects Analysis for REMA Outcome -By site: Time 2 (Series 1)

Fixed Effect									
Effect	Site 1: NY			Site 2: MA			Site 3: TN		
	b	SE	p	b	SE	p	b	SE	p
Intercept	15.11	6.14	0.0179	22.80	6.38	0.0008	36.00	6.45	0.0000
Condition	3.07	0.54	0.0000	1.44	0.74	0.0684	3.19	0.45	0.0000
REMA pretest school-mean centered	0.34	0.02	0.0000	0.40	0.04	0.0000	0.59	0.04	0.0000
REMA pretest school mean	0.43	0.14	0.0046	0.36	0.12	0.0079	0.17	0.13	0.1982
Age at Time 2	0.20	0.03	0.0000	0.15	0.05	0.0070	0.10	0.05	0.0648
Age interval between Time 2 and Time 1	-0.17	0.26	0.5179	-0.06	0.33	0.8595	0.42	0.50	0.4020
Test Lag from School Start to Pretest	-0.03	0.27	0.9014	-0.21	0.53	0.6975	-0.54	0.54	0.3180
Ethnicity: Black	-1.23	0.63	0.0520	-0.20	0.82	0.8049	-1.61	1.04	0.1224
Ethnicity: White	0.83	0.68	0.2222	0.82	0.95	0.3836	-0.50	1.22	0.6844
Ethnicity: Hispanic	-0.37	0.70	0.5929	-0.05	0.78	0.9509	-0.82	1.58	0.6039
English Language Learner Status	-0.26	0.61	0.6648	1.43	0.50	0.0044	0.85	1.64	0.6050
Gender (Male)	-0.48	0.24	0.0442	-0.50	0.41	0.2266	-0.40	0.39	0.3159
Parent Education	0.50	0.14	0.0003	0.77	0.24	0.0014	0.28	0.23	0.2284

Random Effect									
Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	0.87	0.39	0.0124	0.00			0.00		
Level 3 (School)	0.57	0.37	0.0637	1.08	0.72	0.0673	0.17	0.29	0.2853
Level 1 (Student)	12.89	0.62	0.0000	14.38	1.12	0.0000	14.72	1.06	0.0000

Random Effect - Unconditional Model									
Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	0.77	0.48	0.053099	0.00			1.19	1.20	0.159745
Level 3 (School)	3.73	1.32	0.00228	2.04	1.19	0.043815	1.93	1.41	0.085556
Level 1 (Student)	19.08	0.92	1.28E-96	22.44	1.72	3.69E-39	26.68	1.95	5.84E-43

Table A11: Fixed and Random Effects Analysis for REMA Outcome -By site: Time 3 (Series 1)

Fixed Effect

Effect	Site 1: NY			Site 2: MA			Site 3: TN		
	b	SE	p	b	SE	p	b	SE	p
Intercept	28.93	9.27	0.0024	49.80	11.63	0.0000	31.29	9.12	0.0008
Condition	0.87	0.54	0.1173	0.91	0.78	0.2766	0.62	0.38	0.1093
REMA pretest school-mean centered	0.33	0.03	0.0000	0.33	0.05	0.0000	0.50	0.04	0.0000
REMA pretest school mean	0.36	0.14	0.0243	0.40	0.11	0.0062	0.35	0.10	0.0015
Age at Time 3	0.17	0.04	0.0001	0.04	0.07	0.5736	0.07	0.05	0.1776
Age interval between Time 3 and Time 1	0.05	0.35	0.8853	0.67	0.51	0.1861	-0.18	0.38	0.6408
Test Lag from School Start to Pretest	-0.47	0.35	0.1897	-0.29	0.77	0.7078	0.25	0.50	0.6186
Ethnicity: Black	-2.66	0.80	0.0010	-3.74	1.13	0.0011	-1.56	1.06	0.1413
Ethnicity: White	0.08	0.85	0.9256	-2.87	1.30	0.0285	0.36	1.19	0.7600
Ethnicity: Hispanic	-1.38	0.86	0.1113	-2.90	1.13	0.0110	-1.63	1.51	0.2796
English Language Learner Status	0.54	0.69	0.4389	1.61	0.66	0.0151	0.74	1.58	0.6389
Gender (Male)	0.07	0.30	0.8110	-0.95	0.52	0.0707	-0.21	0.37	0.5622
Parent Education	0.42	0.18	0.0185	1.46	0.31	0.0000	-0.03	0.22	0.8789

Random Effect

Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	0.77	0.47	0.0508	0.00			0.04	0.27	0.4412
Level 3 (School)	0.15	0.33	0.3217	0.60	0.86	0.2418	0.00		
Level 1 (Student)	12.66	0.79	0.0000	12.66	1.35	0.0000	13.11	0.96	0.0000

Random Effect - Unconditional Model

Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	1.14	0.69	0.0490	0.00			0.32	0.49	0.2579
Level 3 (School)	1.71	0.92	0.0316	3.39	2.10	0.0531	0.00		
Level 1 (Student)	18.50	1.14	0.0000	18.99	1.96	0.0000	21.74	1.57	0.0000

Table A12: Fixed and Random Effects Analysis for REMA Outcome -By site: Time 4 (Series 1)

Fixed Effect									
Effect	Site 1: NY			Site 2: MA			Site 3: TN		
	b	SE	p	b	SE	p	b	SE	p
Intercept	49.73	12.41	0.0001	88.84	20.22	0.0000	47.46	9.73	0.0000
Condition	0.55	0.60	0.3686	0.55	1.00	0.5958	0.75	0.41	0.0679
REMA pretest school-mean centered	0.33	0.03	0.0000	0.33	0.06	0.0000	0.52	0.04	0.0000
REMA pretest school mean	0.32	0.17	0.0750	0.58	0.15	0.0044	0.27	0.11	0.0134
Age at Time 4	0.09	0.05	0.0526	-0.16	0.08	0.0358	0.04	0.05	0.4398
Age interval between Time 4 and Time 1	0.27	0.35	0.4390	1.17	0.58	0.0436	0.07	0.25	0.7773
Test Lag from School Start to Pretest	-0.16	0.40	0.6923	-0.40	0.86	0.6405	0.11	0.47	0.8150
Ethnicity: Black	-3.11	0.91	0.0007	-3.24	1.28	0.0126	-1.55	1.14	0.1731
Ethnicity: White	-0.26	0.97	0.7903	-0.94	1.43	0.5102	0.19	1.28	0.8852
Ethnicity: Hispanic	-1.61	0.99	0.1047	-0.91	1.26	0.4720	-1.54	1.62	0.3415
English Language Learner Status	0.20	0.78	0.7990	1.33	0.71	0.0619	0.49	1.71	0.7740
Gender (Male)	0.40	0.34	0.2442	0.51	0.56	0.3724	0.03	0.40	0.9433
Parent Education	0.66	0.20	0.0009	1.37	0.31	0.0000	-0.11	0.23	0.6373

Random Effect									
Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	0.83	0.56	0.0686	0.45	1.28	0.3632	0.00		
Level 3 (School)	0.32	0.47	0.2513	1.18	1.73	0.2469	0.00		
Level 1 (Student)	14.32	0.93	0.0000	13.37	1.51	0.0000	15.53	1.10	0.0000

Random Effect - Unconditional Model									
Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	0.91	0.72	0.1019	1.23	1.98	0.2680	0.00		
Level 3 (School)	1.70	0.95	0.0364	4.12	3.33	0.1081	0.00		
Level 1 (Student)	20.09	1.29	0.0000	19.16	2.08	0.0000	24.69	1.73	0.0000

Table A13: Summary of Moderator Analyses for REMA Outcome: Time 2 (Series 1)

Moderator	Male		ELL		Pretest		Age_T2		Black		Hispanic		White	
	b	p	b	p	b	p	b	p	b	p	b	p	b	p
Intercept	44.21	0.0000	44.20	0.0000	44.30	0.0000	44.30	0.0000	44.63	0.0000	44.01	0.0000	44.15	0.0000
Condition	2.65	0.0000	2.75	0.0000	2.81	0.0000	2.82	0.0000	2.07	0.0000	2.95	0.0000	2.85	0.0000
Condition*Moderator	0.24	0.5454	0.06	0.9148	-0.13	0.0002	-0.13	0.0107	1.23	0.0073	-0.91	0.1054	-0.52	0.3764
Site 2: MA	1.49	0.0016	1.48	0.0018	1.42	0.0022	1.51	0.0012	1.53	0.0009	1.52	0.0012	1.49	0.0016
Site 1: NY	0.72	0.0648	0.72	0.0649	0.68	0.0759	0.75	0.0512	0.69	0.0671	0.70	0.0690	0.73	0.0607
Site 3: TN	0		0		0		0		0		0		0	
REMA pretest grand-mean centered	0.41	0.0000	0.41	0.0000	0.50	0.0000	0.41	0.0000	0.42	0.0000	0.41	0.0000	0.41	0.0000
Age at Time 2	0.16	0.0000	0.16	0.0000	0.16	0.0000	0.24	0.0000	0.16	0.0000	0.16	0.0000	0.16	0.0000
Age interval between Time 2 and Time 1	-0.07	0.6912	-0.07	0.7197	-0.05	0.7838	-0.04	0.8122	-0.07	0.6950	-0.08	0.6840	-0.07	0.7267
Test Lag from School Start to Pretest	-0.17	0.4456	-0.17	0.4330	-0.16	0.4591	-0.15	0.5036	-0.19	0.3887	-0.18	0.4138	-0.18	0.4192
Ethnicity: Black	-1.16	0.0087	-1.17	0.0084	-1.17	0.0083	-1.19	0.0070	-1.96	0.0002	-1.16	0.0088	-1.17	0.0084
Ethnicity: White	0.57	0.2429	0.57	0.2457	0.61	0.2082	0.54	0.2676	0.57	0.2437	0.57	0.2406	0.90	0.1445
Ethnicity: Hispanic	-0.46	0.3222	-0.47	0.3183	-0.46	0.3232	-0.49	0.2942	-0.48	0.3051	0.12	0.8413	-0.46	0.3214
English Language Learner Status	0.68	0.0548	0.64	0.1958	0.74	0.0360	0.66	0.0596	0.61	0.0835	0.63	0.0761	0.67	0.0562
Gender (Male)	-0.66	0.0362	-0.51	0.0060	-0.53	0.0044	-0.50	0.0070	-0.50	0.0063	-0.51	0.0058	-0.51	0.0061
Parent Education	0.48	0.0000	0.48	0.0000	0.48	0.0000	0.48	0.0000	0.48	0.0000	0.48	0.0000	0.48	0.0000

Table A14: Summary of Moderator Analyses for REMA Outcome: Time 3 (Series 1)

Moderator	Male		ELL		Pretest		Age_T3		Black		Hispanic		White	
	b	p	b	p	b	p	b	p	b	p	b	p	b	p
Intercept	54.49	0.0000	54.41	0.0000	54.75	0.0000	54.49	0.0000	54.57	0.0000	54.46	0.0000	54.50	0.0000
Condition	0.81	0.0177	0.78	0.0085	0.75	0.0080	0.76	0.0068	0.39	0.3246	0.77	0.0108	0.73	0.0143
Condition*Moderator	-0.16	0.7144	-0.30	0.6488	-0.04	0.2875	-0.08	0.1517	0.56	0.2394	-0.16	0.7986	0.06	0.9264
Site 2: MA	1.84	0.0000	1.86	0.0000	1.82	0.0000	1.88	0.0000	1.88	0.0000	1.85	0.0000	1.84	0.0000
Site 1: NY	0.11	0.7059	0.10	0.7279	0.10	0.7321	0.12	0.6708	0.12	0.6875	0.11	0.7141	0.11	0.7192
Site 3: TN	0		0		0		0		0		0		0	
REMA pretest grand-mean centered	0.39	0.0000	0.39	0.0000	0.41	0.0000	0.39	0.0000	0.39	0.0000	0.39	0.0000	0.39	0.0000
Age at Time 3	0.11	0.0001	0.11	0.0001	0.11	0.0001	0.15	0.0002	0.11	0.0001	0.11	0.0001	0.11	0.0001
Age interval between Time 3 and Time 1	0.04	0.8630	0.03	0.8791	0.05	0.8206	0.04	0.8565	0.03	0.8907	0.04	0.8727	0.04	0.8677
Test Lag from School Start to Pretest	-0.32	0.2360	-0.31	0.2418	-0.31	0.2401	-0.28	0.3026	-0.33	0.2108	-0.31	0.2411	-0.31	0.2467
Ethnicity: Black	-2.69	0.0000	-2.68	0.0000	-2.67	0.0000	-2.68	0.0000	-2.96	0.0000	-2.69	0.0000	-2.69	0.0000
Ethnicity: White	-0.45	0.4412	-0.44	0.4530	-0.41	0.4786	-0.46	0.4280	-0.41	0.4785	-0.44	0.4449	-0.47	0.4797
Ethnicity: Hispanic	-1.75	0.0022	-1.75	0.0023	-1.73	0.0024	-1.75	0.0021	-1.75	0.0021	-1.68	0.0080	-1.75	0.0022
English Language Learner Status	0.75	0.0716	0.88	0.0763	0.79	0.0611	0.77	0.0669	0.73	0.0808	0.75	0.0741	0.75	0.0719
Gender (Male)	-0.14	0.6433	-0.23	0.2937	-0.23	0.2797	-0.22	0.3169	-0.23	0.2814	-0.23	0.2956	-0.23	0.2959
Parent Education	0.46	0.0003	0.46	0.0003	0.45	0.0004	0.45	0.0003	0.44	0.0004	0.45	0.0003	0.46	0.0003

Table A15: Summary of Moderator Analyses for REMA Outcome: Time 4 (Series 1)

Moderator	Male		ELL		Pretest		Age_T4		Black		Hispanic		White	
	b	p	b	p	b	p	b	p	b	p	b	p	b	p
Intercept	64.49	0.0000	64.34	0.0000	64.45	0.0000	64.46	0.0000	64.21	0.0000	64.38	0.0000	64.50	0.0000
Condition	0.62	0.1066	0.56	0.0934	0.51	0.1051	0.51	0.1066	0.07	0.8727	0.58	0.0806	0.50	0.1289
Condition*Moderator	-0.23	0.6231	-0.32	0.6605	0.00	0.9376	0.01	0.8597	0.69	0.1915	-0.43	0.5273	0.03	0.9622
Site 2: MA	3.05	0.0000	3.07	0.0000	3.06	0.0000	3.05	0.0000	3.07	0.0000	3.07	0.0000	3.05	0.0000
Site 1: NY	1.57	0.0001	1.56	0.0001	1.57	0.0001	1.57	0.0001	1.57	0.0001	1.57	0.0001	1.57	0.0001
Site 3: TN	0		0		0		0		0		0		0	
REMA pretest grand-mean centered	0.40	0.0000	0.40	0.0000	0.40	0.0000	0.40	0.0000	0.40	0.0000	0.40	0.0000	0.40	0.0000
Age at Time 4	0.04	0.1787	0.04	0.1835	0.04	0.1844	0.04	0.4077	0.04	0.1918	0.04	0.1820	0.04	0.1846
Age interval between Time 4 and Time 1	0.16	0.4056	0.16	0.4212	0.16	0.4138	0.16	0.4133	0.14	0.4606	0.16	0.4173	0.16	0.4096
Test Lag from School Start to Pretest	-0.11	0.7003	-0.10	0.7207	-0.10	0.7225	-0.10	0.7142	-0.12	0.6592	-0.11	0.7084	-0.10	0.7253
Ethnicity: Black	-2.84	0.0000	-2.83	0.0000	-2.84	0.0000	-2.84	0.0000	-3.17	0.0000	-2.84	0.0000	-2.84	0.0000
Ethnicity: White	-0.48	0.4576	-0.47	0.4696	-0.47	0.4625	-0.47	0.4660	-0.44	0.4938	-0.48	0.4593	-0.49	0.5067
Ethnicity: Hispanic	-1.53	0.0158	-1.53	0.0160	-1.54	0.0155	-1.54	0.0156	-1.54	0.0154	-1.34	0.0570	-1.54	0.0154
English Language Learner Status	0.57	0.2249	0.70	0.2035	0.57	0.2219	0.57	0.2204	0.55	0.2370	0.56	0.2277	0.57	0.2205
Gender (Male)	0.33	0.3380	0.21	0.3861	0.21	0.3823	0.21	0.3862	0.20	0.4034	0.21	0.3787	0.21	0.3820
Parent Education	0.51	0.0002	0.51	0.0002	0.51	0.0002	0.51	0.0002	0.50	0.0002	0.51	0.0002	0.51	0.0002

Table A16: 3-way Interaction of Pretest by Condition by Site for REMA Outcome (Complete Sample)

Effect	p-value Time 2	p-value Time 3	p-value Time 4
Condition	0.0000	0.0104	0.1291
Condition*Pretest	0.0050	0.3472	0.9353
Condition*Site*Pretest	0.0078	0.0260	0.0052
Site	0.0160	0.0001	0.0000
Pretest	0.0000	0.0000	0.0000
Age at Assessment Time	0.0000	0.0002	0.2480
Age interval between Assessment Time and Time 1	0.7678	0.9610	0.5315
Test Lag from School Start to Pretest	0.4359	0.2279	0.6812
Ethnicity: Black	0.0121	0.0000	0.0000
Ethnicity: White	0.1448	0.5975	0.5594
Ethnicity: Hispanic	0.4122	0.0031	0.0168
English Language Learner Status	0.0264	0.0558	0.2154
Gender (Male)	0.0056	0.2935	0.3820
Parent Education	0.0000	0.0004	0.0001

Note: Entries are *p*-values of significance test conducted by three-level HLM. Pretest was grand-mean centered.

Table A17: Pretest (site-mean centering) as a Moderator by Site for REMA Outcome (Complete Sample: Time 2)

Effect	Buffalo			Boston			Nashville		
	b	SE	p	b	SE	p	b	SE	p
Intercept	43.86	2.15	0.0000	45.75	2.81	0.0000	51.54	4.22	0.0000
Condition	3.04	0.53	0.0000	1.46	0.73	0.0609	3.10	0.53	0.0000
Condition*Pretest	-0.09	0.05	0.0528	0.00	0.08	0.9986	-0.18	0.07	0.0101
REMA Pretest (site-mean centering)	0.41	0.04	0.0000	0.40	0.07	0.0000	0.66	0.05	0.0000
Age at Time 2	0.20	0.03	0.0000	0.15	0.05	0.0067	0.10	0.05	0.0562
Age interval between Time 2 and Time 1	-0.16	0.26	0.5301	-0.06	0.33	0.8432	0.70	0.52	0.1804
Test Lag from School Start to Pretest	-0.04	0.27	0.8744	-0.21	0.53	0.6974	-0.77	0.54	0.1544
Ethnicity: Black	-1.23	0.63	0.0518	-0.17	0.82	0.8327	-2.29	1.04	0.0281
Ethnicity: White	0.90	0.68	0.1812	0.83	0.95	0.3792	-1.08	1.22	0.3771
Ethnicity: Hispanic	-0.37	0.70	0.5963	-0.01	0.77	0.9911	-0.96	1.59	0.5476
English Language Learner Status	-0.17	0.61	0.7766	1.44	0.50	0.0040	0.36	1.64	0.8250
Gender (Male)	-0.51	0.24	0.0350	-0.50	0.42	0.2267	-0.50	0.39	0.2058
Parent Education	0.51	0.14	0.0002	0.76	0.24	0.0015	0.18	0.23	0.4437

Table A18: Pretest (site-mean centering) as a Moderator by Site for REMA Outcome (Complete Sample: Time 3)

Effect	Buffalo			Boston			Nashville		
	b	SE	p	b	SE	p	b	SE	p
Intercept	54.91	6.94	0.0000	68.70	10.00	0.0000	49.33	7.83	0.0000
Condition	0.86	0.52	0.1053	0.93	0.75	0.2557	0.61	0.38	0.1201
Condition*Pretest	-0.05	0.05	0.3108	-0.02	0.09	0.8637	0.00	0.06	0.9650
REMA Pretest (site-mean centering)	0.37	0.04	0.0000	0.35	0.07	0.0000	0.48	0.05	0.0000
Age at Time 3	0.17	0.04	0.0001	0.04	0.07	0.5889	0.07	0.05	0.1714
Age interval between Time 3 and Time 1	0.06	0.35	0.8690	0.69	0.51	0.1790	-0.20	0.38	0.5982
Test Lag from School Start to Pretest	-0.48	0.35	0.1779	-0.25	0.78	0.7488	0.15	0.50	0.7650
Ethnicity: Black	-2.66	0.80	0.0010	-3.75	1.13	0.0011	-1.83	1.04	0.0793
Ethnicity: White	0.12	0.85	0.8921	-2.86	1.30	0.0294	0.17	1.19	0.8843
Ethnicity: Hispanic	-1.37	0.86	0.1118	-3.01	1.12	0.0080	-1.65	1.51	0.2742
English Language Learner Status	0.63	0.68	0.3594	1.63	0.66	0.0143	0.59	1.58	0.7114
Gender (Male)	0.06	0.30	0.8493	-0.96	0.53	0.0701	-0.27	0.37	0.4637
Parent Education	0.42	0.18	0.0181	1.48	0.31	0.0000	-0.07	0.22	0.7515

Table A19: Pretest (site-mean centering) as a Moderator by Site for REMA Outcome (Complete Sample: Time 4)

Effect	Buffalo			Boston			Nashville		
	b	SE	p	b	SE	p	b	SE	p
Intercept	69.47	11.14	0.0000	96.80	18.28	0.0000	59.08	8.21	0.0000
Condition	0.56	0.58	0.3436	0.48	1.06	0.6631	0.65	0.42	0.1455
Condition*Pretest	0.00	0.06	0.9955	-0.05	0.11	0.6203	0.08	0.07	0.2783
REMA Pretest (site-mean centering)	0.33	0.05	0.0000	0.38	0.09	0.0000	0.46	0.05	0.0000
Age at Time 4	0.09	0.05	0.0532	-0.17	0.08	0.0314	0.04	0.06	0.4273
Age interval between Time 4 and Time 1	0.27	0.35	0.4433	1.13	0.58	0.0538	-0.02	0.26	0.9402
Test Lag from School Start to Pretest	-0.16	0.40	0.6931	-0.30	0.87	0.7340	-0.12	0.48	0.7993
Ethnicity: Black	-3.11	0.91	0.0006	-3.36	1.30	0.0104	-2.00	1.13	0.0790
Ethnicity: White	-0.26	0.96	0.7834	-0.83	1.44	0.5683	-0.13	1.30	0.9203
Ethnicity: Hispanic	-1.58	0.99	0.1099	-1.14	1.27	0.3687	-1.64	1.64	0.3185
English Language Learner Status	0.23	0.77	0.7651	1.31	0.71	0.0682	0.31	1.72	0.8587
Gender (Male)	0.39	0.34	0.2469	0.52	0.57	0.3681	-0.09	0.40	0.8265
Parent Education	0.66	0.20	0.0009	1.40	0.31	0.0000	-0.13	0.23	0.5856

Table A20: *p*-values from Testing Differences on the COEMET Measures across Site and Condition

Effect	Classroom Culture	Number SMAs	Quality SMAs	Number Mini SMAs	COEMET Composite
Site	0.3447	0.0005	0.0000	0.0007	0.6025
Condition	0.0000	0.0000	0.0000	0.0002	<.0001
Condition*Site	0.4725	0.9302	0.0173	0.0135	0.3709

Note: Two-level HLM was used for analysis with the COEMET scores as dependent variables. Entries are *p*-values.

Sample size:

Site	N1 (class)	N2 (school)
Buffalo	75	25
Boston	31	18
Nashville	33	16
Total	139	59

Table A21: Detailed Results of Testing Differences on the COEMET Measures across Site and Condition

Effect	Classroom Culture			Number SMAs			Quality SMAs			Number Mini SMAs			COEMET Composite		
	b	SE	p	b	SE	p	b	SE	p	b	SE	p	b	SE	p
Intercept	3.28	0.13	0.0000	1.74	0.30	0.0000	3.23	0.06	0.0000	5.36	0.62	0.0000	-0.64	0.17	0.0003
Site Boston	0.04	0.21	0.8410	1.92	0.48	0.0002	-0.02	0.11	0.8427	-3.19	0.97	0.0017	-0.26	0.28	0.3548
Site Buffalo	0.25	0.18	0.1738	0.32	0.40	0.4161	0.53	0.08	0.0000	-3.05	0.82	0.0005	-0.02	0.22	0.9332
Site Nashville	0.00			0.00			0.00			0.00			0.00		
Condition	0.82	0.19	0.0001	1.22	0.43	0.0060	0.40	0.09	0.0001	3.92	0.88	0.0000	1.32	0.24	<.0001
Condition*Site Boston	-0.11	0.27	0.6824	-0.08	0.63	0.9017	0.04	0.14	0.7560	-2.67	1.27	0.0397	0.01	0.36	0.9808
Condition*Site Buffalo	-0.29	0.24	0.2357	0.13	0.54	0.8154	-0.26	0.11	0.0257	-3.35	1.12	0.0041	-0.35	0.30	0.2443
Condition*Site Nashville	0.00			0.00			0.00			0.00			0.00		

The COEMET variables are the dependent variables in these analyses.

Table A22: Factor Loadings for the COEMET Subscales (classroom-level analysis)

COEMET Measure (Site-mean Centered)	Loading
Classroom Culture	0.85
Number of SMAs	0.68
Quality of SMAs	0.57
Number Mini SMAs	0.50

N = 139. Principal Factor analysis was conducted.

Table A23: COEMET Composite Variable Predicting Pre-K REMA Gain (Series 1: Time 2)

Effect	b	SE	p
Intercept	44.42	1.58	0.0000
Site 2: MA	2.02	0.50	0.0001
Site 1: NY	1.16	0.43	0.0095
Site 3: TN	0.00		
REMA pretest grand-mean centered	0.41	0.02	0.0000
Age at Time 2	0.16	0.03	0.0000
Age interval between Time 2 and Time 1	-0.22	0.19	0.2316
Test Lag from School Start to Pretest	-0.25	0.22	0.2511
Ethnicity: Black	-1.09	0.45	0.0145
Ethnicity: White	0.62	0.49	0.2072
Ethnicity: Hispanic	-0.39	0.47	0.4039
English Language Learner Status	0.63	0.35	0.0769
Gender (Male)	-0.53	0.19	0.0041
Parent Education	0.49	0.11	0.0000
COEMET composite	1.29	0.32	0.0001
COEMET composite*site 2	-0.41	0.49	0.3983
COEMET composite*site 1	-0.20	0.40	0.6152
COEMET composite*site 3	0.00		

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	p
Site	2.00	60.60	8.2900	0.0007
REMA pretest grand-mean centered	1.00	1692.00	535.4700	<.0001
Age at Time 2	1.00	1661.00	36.8500	<.0001
Age interval between Time 2 and Time 1	1.00	1577.00	1.4300	0.2316
Test Lag from School Start to Pretest	1.00	1686.00	1.3200	0.2511
Ethnicity: Black	1.00	1684.00	5.9900	0.0145
Ethnicity: White	1.00	1661.00	1.5900	0.2072
Ethnicity: Hispanic	1.00	1664.00	0.7000	0.4039
English Language Learner Status	1.00	1665.00	3.1300	0.0769
Gender (Male)	1.00	1643.00	8.2700	0.0041
Parent Education	1.00	1690.00	21.0200	<.0001
COEMET composite	1.00	111.00	35.1100	<.0001
COEMET composite*site	2.00	104.00	0.3600	0.6977

Table A24: COEMET Site-Centered Subscales Predicting REMA Pre-K Gain (Series 1: Time 2)

Effect	b	SE	p
Intercept	44.59	1.58	0.0000
Site 2: MA	2.03	0.51	0.0002
Site 1: NY	1.16	0.44	0.0112
Site 3: TN	0.00		
REMA pretest grand-mean centered	0.41	0.02	0.0000
Age at Time 2	0.16	0.03	0.0000
Age interval between Time 2 and Time 1	-0.20	0.19	0.2888
Test Lag from School Start to Pretest	-0.25	0.22	0.2646
Ethnicity: Black	-1.09	0.45	0.0149
Ethnicity: White	0.61	0.49	0.2117
Ethnicity: Hispanic	-0.37	0.47	0.4287
English Language Learner Status	0.65	0.35	0.0650
Gender (Male)	-0.52	0.19	0.0049
Parent Education	0.49	0.11	0.0000
Classroom Culture	0.39	0.22	0.0807
Number of SMAs	0.24	0.17	0.1702
Quality of SMAs	0.38	0.16	0.0190
Number of Mini SMAs	0.27	0.16	0.0897

Table A25: COEMET Composite Variable Predicting REMA Gain through K
(Series 1: Time 3)

Effect	b	SE	p
Intercept	53.69	4.43	0.0000
Site 2: MA	1.95	0.40	0.0000
Site 1: NY	0.28	0.30	0.3569
Site 3: TN	0.00		
REMA pretest grand-mean centered	0.39	0.02	0.0000
Age at Time 3	0.11	0.03	0.0002
Age interval between Time 3 and Time 1	-0.02	0.22	0.9351
Test Lag from School Start to Pretest	-0.31	0.27	0.2477
Ethnicity: Black	-2.57	0.54	0.0000
Ethnicity: White	-0.44	0.58	0.4488
Ethnicity: Hispanic	-1.68	0.57	0.0034
English Language Learner Status	0.76	0.42	0.0696
Gender (Male)	-0.23	0.22	0.2800
Parent Education	0.47	0.13	0.0002
COEMET composite	0.16	0.22	0.4640
COEMET composite*site 2	0.09	0.38	0.8068
COEMET composite*site 1	0.56	0.31	0.0771
COEMET composite*site 3	0.00		

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	p
Site	2.00	53.10	12.3200	<.0001
REMA pretest grand-mean centered	1.00	1122.00	381.3600	<.0001
Age at Time 3	1.00	1167.00	13.8900	0.0002
Age interval between Time 3 and Time 1	1.00	715.00	0.0100	0.9351
Test Lag from School Start to Pretest	1.00	1053.00	1.3400	0.2477
Ethnicity: Black	1.00	1128.00	22.5100	<.0001
Ethnicity: White	1.00	988.00	0.5700	0.4488
Ethnicity: Hispanic	1.00	998.00	8.6400	0.0034
English Language Learner Status	1.00	1032.00	3.3000	0.0696
Gender (Male)	1.00	1164.00	1.1700	0.2800
Parent Education	1.00	1113.00	13.8100	0.0002
COEMET composite	1.00	79.00	6.5100	0.0126
COEMET composite*site	2.00	71.60	1.7800	0.1765

Table A26: COEMET Site-Centered Subscales Predicting REMA Gain through K
(Series 1: Time 3)

Effect	b	SE	p
Intercept	52.98	4.46	0.0000
Site 2: MA	1.91	0.40	0.0000
Site 1: NY	0.25	0.30	0.4168
Site 3: TN	0.00		
REMA pretest grand-mean centered	0.39	0.02	0.0000
Age at Time 3	0.11	0.03	0.0002
Age interval between Time 3 and Time 1	-0.05	0.22	0.8181
Test Lag from School Start to Pretest	-0.34	0.27	0.2028
Ethnicity: Black	-2.53	0.54	0.0000
Ethnicity: White	-0.36	0.58	0.5366
Ethnicity: Hispanic	-1.56	0.57	0.0068
English Language Learner Status	0.77	0.42	0.0674
Gender (Male)	-0.20	0.22	0.3548
Parent Education	0.48	0.13	0.0001
Classroom Culture	0.04	0.20	0.8396
Number of SMAs	0.25	0.19	0.1878
Quality of SMAs	0.29	0.14	0.0407
Number of Mini SMAs	-0.13	0.16	0.4215

Table A27: COEMET Composite Variable Predicting REMA Gain through 1st Grade (Series 1: Time 4)

Effect	b	SE	p
Intercept	64.66	6.03	<.0001
Site 2: MA	3.12	0.46	<.0001
Site 1: NY	1.73	0.39	<.0001
Site 3: TN	0.00	.	.
REMA pretest grand-mean centered	0.40	0.02	<.0001
Age at Time 4	0.04	0.03	0.2067
Age interval between Time 4 and Time 1	0.16	0.19	0.4068
Test Lag from School Start to Pretest	-0.09	0.28	0.7374
Ethnicity: Black	-2.73	0.60	<.0001
Ethnicity: White	-0.47	0.65	0.4665
Ethnicity: Hispanic	-1.48	0.64	0.0203
English Language Learner Status	0.59	0.47	0.2096
Gender (Male)	0.20	0.24	0.3870
Parent Education	0.51	0.13	0.0001
COEMET composite	0.22	0.25	0.3852
COEMET composite*site 2	-0.06	0.43	0.8940
COEMET composite*site 1	0.39	0.35	0.2685
COEMET composite*site 3	0.00	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	p
site	2.00	58.90	24.0500	<.0001
REMA pretest grand-mean centered	1.00	1066.00	320.4900	<.0001
Age at Time 4	1.00	1103.00	1.6000	0.2067
Age interval between Time 4 and Time 1	1.00	908.00	0.6900	0.4068
Test Lag from School Start to Pretest	1.00	774.00	0.1100	0.7374
Ethnicity: Black	1.00	1056.00	20.6500	<.0001
Ethnicity: White	1.00	942.00	0.5300	0.4665
Ethnicity: Hispanic	1.00	950.00	5.4000	0.0203
English Language Learner Status	1.00	1018.00	1.5800	0.2096
Gender (Male)	1.00	1096.00	0.7500	0.3870
Parent Education	1.00	1066.00	14.5400	0.0001
COEMET composite	1.00	74.10	3.8500	0.0534
COEMET composite*site	2.00	64.90	0.8400	0.4346

Table A28: COEMET Site-Centered Subscales Predicting REMA Gain through 1st Grade

(Series 1: Time 4)

Effect	b	SE	p
Intercept	64.04	6.01	0.0000
Site 2: MA	3.08	0.45	0.0000
Site 1: NY	1.64	0.38	0.0001
Site 3: TN	0.00		
REMA pretest grand-mean centered	0.39	0.02	0.0000
Age at Time 4	0.04	0.03	0.2127
Age interval between Time 4 and Time 1	0.14	0.19	0.4702
Test Lag from School Start to Pretest	-0.15	0.28	0.5902
Ethnicity: Black	-2.73	0.60	0.0000
Ethnicity: White	-0.43	0.64	0.5009
Ethnicity: Hispanic	-1.37	0.64	0.0318
English Language Learner Status	0.59	0.47	0.2052
Gender (Male)	0.23	0.24	0.3353
Parent Education	0.51	0.13	0.0002
Classroom Culture	0.26	0.22	0.2372
Number of SMAs	0.09	0.21	0.6652
Quality of SMAs	0.19	0.16	0.2259
Number of Mini SMAs	-0.26	0.18	0.1490

Table A29: The COEMET Composite Variable as a Moderator of the Effects of Condition on REMA Gain During Pre-K (Series 1: Time 2)

Effect	b	SE	p
Intercept	44.50	1.56	<.0001
REMA Pretest	0.41	0.02	<.0001
Age at Time 2	0.16	0.03	<.0001
Age interval between Time 2 and Time 1	-0.13	0.19	0.4928
Test Lag from School Start to Pretest	-0.15	0.22	0.4896
Ethnicity: Black	-1.23	0.44	0.0054
Ethnicity: White	0.59	0.49	0.2289
Ethnicity: Hispanic	-0.32	0.47	0.4864
English Language Learner Status	0.80	0.35	0.0226
Gender (Male)	-0.51	0.18	0.0061
Parent Education	0.51	0.11	<.0001
Condition	2.38	0.42	<.0001
COEMET Composite	0.26	0.29	0.3686
Condition*COEMET Composite	0.49	0.39	0.2044

Table A30: Analysis of the COEMET Composite Variable as a Mediator the Effects of Condition on the REMA During Pre-K. Part 1: Sample Size and Model Fit Statistics (Series 1: Time 2)

Sample Size	
N (classroom)	139
J (school)	59
Model Fit Statistics	
Chi-square	47.210
p-value	0.000
RMSEA (Root Mean Square Error Of Approximation)	0.164
CFI	0.814
TLI	0.573
SRMR (Standardized Root Mean Square Residual)	
	Within 0.071
	Between 0.183

Table A31: Analysis of the COEMET Composite Variable as a Mediator the Effects of Condition on REMA During Pre-K. Part 2: Level 1 Within Classroom Results (Series 1: Time 2)

Within Classroom Level Analysis				
REMA2 predicted by ...				
	Estimate	S.E.	Est./S.E.	Two-Tailed <i>p</i> -Value
REMA1	0.432	0.074	5.867	0.000
AGE2	-0.171	0.101	-1.681	0.093
INTERVAL	2.107	0.562	3.753	0.000
TESTLAG	-3.157	0.741	-4.257	0.000
BLACK	-2.888	1.188	-2.432	0.015
WHITE	-3.045	1.329	-2.291	0.022
HISPANIC	-0.041	1.286	-0.032	0.974
MALE	-1.525	0.764	-1.996	0.046
COEMET	0.721	0.196	3.679	0.000
Variances				
COEMET	0.301	0.048	6.304	0.000
Residual				
REMA2	1.615	0.283	5.703	0.000

Table A32: Analysis of the COEMET Composite Variable as a Mediator the Effects of Condition on REMA During Pre-K. Part 3: Level 2 Between School Results (Series 1: Time 2)

Between School Level Analysis				
REMA2 predicted by ...				
	Estimate	S.E.	Est./S.E.	Two-Tailed <i>p</i> -Value
INTERVAL	-5.266	1.252	-4.207	0.000
COEMET	0.847	0.356	2.378	0.017
TREAT	1.373	0.638	2.152	0.031

COEMET predicted by ...				
	Estimate	S.E.	Est./S.E.	Two-Tailed <i>p</i> -Value
TREAT	1.154	0.170	6.797	0.000

	Estimate	S.E.	Est./S.E.	Two-Tailed <i>p</i> -Value
Indirect Effect				
INDB (=0.847*1.154)	0.977	0.380	2.570	0.010

Table A33: Factor Loadings of the Near Fidelity Items Within Setting (item-level analysis)

	Factor 1	Factor 2	Factor 3	Factor 4
General Curriculum				
Q1 within 2 weeks of schedule	.670			
Q2 activities sent home	.650			
Q3 math materials present	.690			
Q4 curriculum used in math activities	.738			
Q5 teacher extended activities	.782			
Hands On Centers				
Q6 children engaged	.872	-.199		
Q7 child selected task	.131	.975		
Q8 materials set up correctly	.771	.134		
Q9 introduced correctly	.874	.050		
Q10 adult monitored as needed	.893	-.163		
Q11 management strategies enhanced quality	.596	.075		
Whole Group				
Q12 teacher understanding	.611			
Q13 materials set up correctly	.833			
Q14 children engaged	.856			
Q15 appropriate pace	.708			
Q16 activity done as written	.656			
Q17 management strategy	.771			
Q18 activity included discussion	.763			
Small Group				
Q19 teacher understanding	.368	.300	.012	.762
Q20 materials set up correctly	.465	.654	.147	.006
Q21 activity done as written	.664	.438	.086	.338
Q22 appropriate pace	.631	.401	-.021	-.153
Q23 activity completed	.177	.245	.680	-.034
Q24 management strategy	.583	.369	-.094	-.201
Q25 valued persistence	.660	-.048	-.289	-.023
Q26 encouraged thinking	.822	-.217	-.188	.129
Q27 asked child to share ideas	.830	-.204	.006	.079
Q28 facilitated response	.769	-.068	-.248	-.154
Q29 encouraged evaluation of others	.754	-.421	.248	-.051
Q30 supported thinking	.792	-.232	-.194	.155
Q31 supported listener's understanding	.683	-.497	.333	.004
Q32 gave just enough assistance	.578	.367	-.323	-.223
Q33 elaborated ideas	.794	-.189	-.338	.108
Q34 went beyond initial solutions	.744	-.085	.135	.156
Q35 encouraged reflection	.723	-.321	.202	-.134
Q36 cultivated love of challenge	.770	-.163	.076	-.183
Q37 completed record sheet	.335	.638	-.012	-.160
Q38 accommodated range of abilities	.707	.169	.002	-.187
Q39 used Monitoring Student Progress	.659	.122	.365	-.097
Computers				

Q40 set up correctly	.617	.615	-.323
Q41 child signed in	.565	.655	-.143
Q42 children engaged	.673	-.278	-.162
Q43 available to help	.829	-.115	-.147
Q44 management strategy	.610	.116	.128
Q45 all will complete activity in week	.588	.512	-.070
Q46 actively involved	.787	-.178	-.292
Q47 strategies appropriate developmentally	.755	-.409	-.142
Q48 realistic expectations	.762	.124	.468
Q49 valued persistence	.700	-.190	.556
Q50 gave just enough assistance	.758	-.461	-.046
Q51 monitored and observed	.797	-.176	-.108
Q52 knows how to access records	.418	.335	.444

Table A34. Zero-order Correlations among Near Fidelity Scores and Classroom Mean Residualized Gain on the REMA

	T2	T3	T4
Option 1			
General	.206	.132	.112
Hands On Centers	-.017	.060	.026
Whole Group	.209	.148	.124
Small Group	.391	.234	.276
Computers	.310	.163	.183
Option 2			
General	.225	.149	.111
Hands On Centers	.059	.103	.054
Whole Group	.230	.190	.155
Small Group	.373	.236	.266
Computers	.318	.147	.177
Option 3			
General	.177	.118	.101
Hands On Centers	-.106	-.041	-.029
Whole Group	.244	.158	.138
Small Group	.258	.130	.192
Computers	.208	.221	.186

Table A35: Fixed and Random Effects of Near Fidelity Additive Composite on REMA Gain: Time 2

Fixed Effects

Effect	b	SE	p
Intercept	24.97	2.69	.000
Site: Buffalo	0.78	0.48	.113
Site: Boston	1.23	0.56	.032
Site: Nashville	0.00		
Male	-0.46	0.22	.037
ELL	1.09	0.45	.015
Black	-0.51	0.53	.343
White	0.95	0.58	.103
Hispanic	-0.43	0.56	.439
Age at Posttest	0.14	0.03	.000
REMA Pretest Score	0.38	0.02	.000
Test Lag from School Start	-0.19	0.26	.470
Interval between Test Dates	0.03	0.26	.895
Highest Parent Education	0.49	0.13	.000
Near Fidelity Additive Score	0.22	0.10	.028
Near Fidelity Additive Score * Site: Buffalo	-0.39	0.30	.206
Near Fidelity Additive Score * Site: Boston	-0.45	0.32	.161
Near Fidelity Additive Score * Site: Nashville	0.00		

Random Effects

Level	Variance	SE	p	ICC
Level 2 (Class)	0.90	0.39	.020	0.06
Level 3 (School)	0.22	0.30	.464	0.02
Level 1 (Student)	12.95	0.57	.000	NA

Table A36: Fixed Effects of Near Fidelity Setting Variables on REMA Gain: Time 2 (N=1138 – Children in All Treatment Classrooms)

Fixed Effects

Effect	General Curric.			Hands On Centers			Whole Group			Small Group			Computers		
	b	SE	p	b	SE	p	b	SE	p	b	SE	p	b	SE	p
Intercept	25.04	2.69	.000	25.24	2.71	.000	25.09	2.70	.000	25.64	2.76	.000	24.72	2.81	.000
Site: Buffalo	1.04	0.50	.043	0.78	0.51	.131	0.85	0.50	.098	0.59	0.52	.262	0.48	0.53	.377
Site: Boston	1.13	0.59	.061	1.40	0.59	.021	1.31	0.58	.029	0.92	0.67	.178	0.84	0.60	.169
Site: Nashville	0.00			0.00			0.00			0.00			0.00		
Male	-0.47	0.22	.035	-0.49	0.22	.028	-0.46	0.22	.036	-0.35	0.22	.122	-0.39	0.23	.084
ELL	1.10	0.44	.013	1.31	0.46	.004	1.01	0.45	.023	1.13	0.46	.015	1.09	0.46	.018
Black	-0.57	0.53	.288	-0.69	0.55	.206	-0.52	0.54	.330	-0.40	0.56	.478	-0.63	0.54	.247
White	0.92	0.58	.117	0.66	0.60	.267	0.93	0.59	.115	1.18	0.61	.053	0.92	0.60	.125
Hispanic	-0.50	0.56	.370	-0.81	0.57	.155	-0.47	0.56	.396	-0.33	0.58	.576	-0.58	0.57	.312
Age at Posttest	0.14	0.03	.000	0.12	0.03	.000	0.14	0.03	.000	0.14	0.03	.000	0.14	0.03	.000
REMA Pretest Score	0.38	0.02	.000	0.39	0.02	.000	0.38	0.02	.000	0.38	0.02	.000	0.38	0.02	.000
Test Lag from School Start	-0.17	0.26	.511	-0.20	0.27	.456	-0.21	0.26	.424	-0.17	0.27	.526	-0.23	0.27	.403
Interval between Test Dates	0.05	0.26	.830	-0.05	0.26	.855	0.02	0.26	.945	0.08	0.26	.746	0.01	0.27	.972
Highest Parent Education	0.48	0.13	.000	0.41	0.13	.001	0.49	0.13	.000	0.46	0.13	.000	0.48	0.13	.000
NF Setting Score	0.50	0.18	.008	-0.03	0.18	.861	0.39	0.20	.050	0.34	0.20	.097	0.37	0.25	.139
NF Setting Score * Site: Buffalo	-0.40	0.52	.445	-0.20	0.46	.667	-1.18	0.64	.070	-0.04	0.46	.929	-0.40	0.60	.499
NF Setting Score * Site: Boston	-0.41	0.60	.493	-0.51	0.54	.348	-1.53	0.71	.034	-0.25	0.61	.688	-0.34	0.59	.567
NF Setting Score * Site: Nashville	0.00			0.00			0.00			0.00			0.00		

Table A37: Fixed Effects of Near Fidelity Setting Variables on REMA Gain: Time 2 (N=992 – Children in Classrooms without Missing Setting Scores)

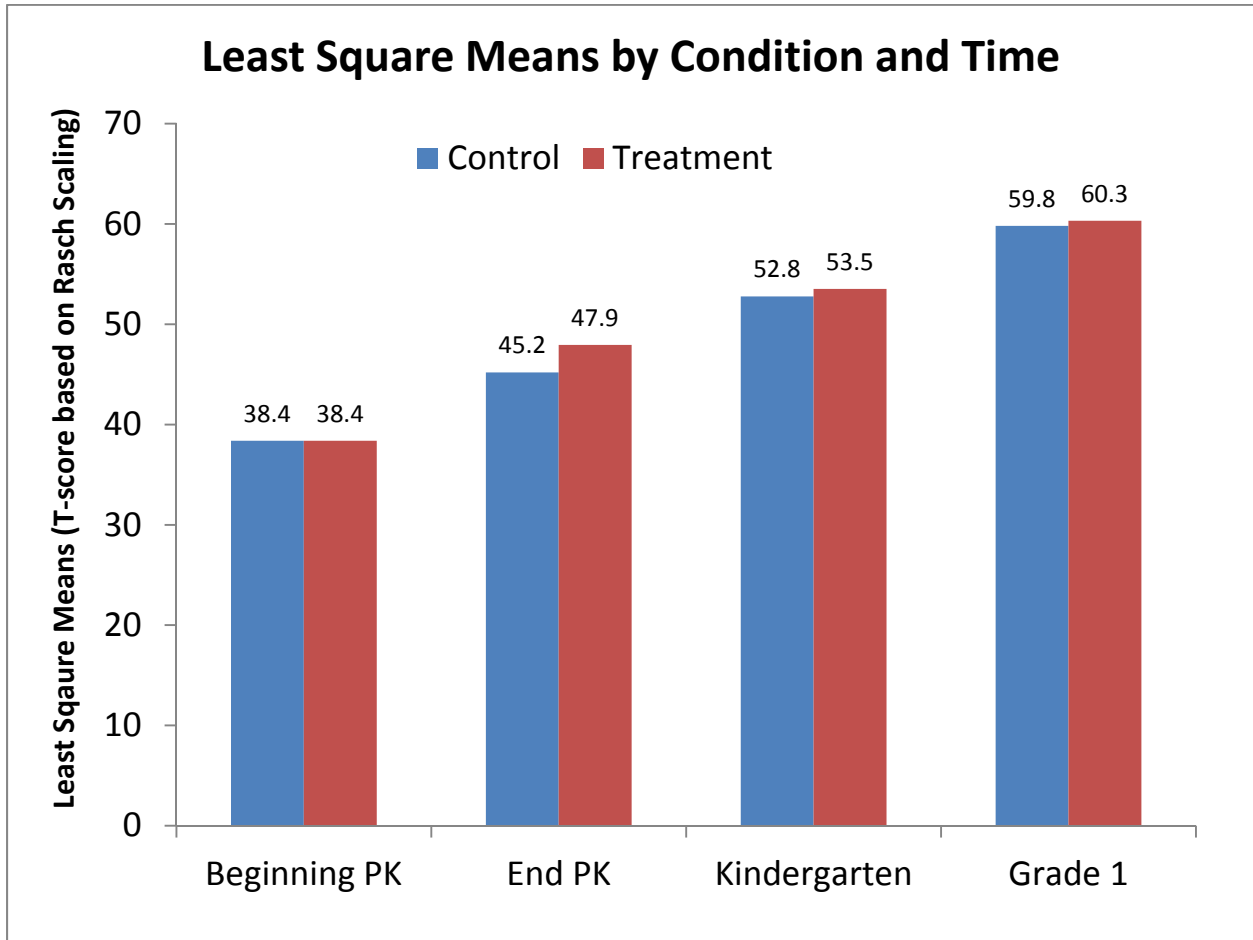
Fixed Effects

Effect	General Curric.			Hands On Centers			Whole Group			Small Group			Computers		
	b	SE	p	b	SE	p	b	SE	p	b	SE	p	b	SE	p
Intercept	25.26	2.88	.000	25.43	2.88	.000	25.15	2.88	.000	25.55	2.89	.000	25.13	2.88	.000
Site: Buffalo	0.92	0.50	.074	0.67	0.51	.197	0.85	0.49	.093	0.60	0.51	.246	0.48	0.51	.354
Site: Boston	1.32	0.64	.042	1.34	0.60	.030	1.47	0.59	.015	1.16	0.70	.104	1.06	0.59	.077
Site: Nashville	0.00			0.00			0.00			0.00			0.00		
Male	-0.30	0.23	.190	-0.30	0.23	.193	-0.30	0.23	.198	-0.29	0.23	.218	-0.29	0.23	.214
ELL	1.49	0.50	.003	1.45	0.49	.003	1.42	0.49	.004	1.47	0.50	.003	1.49	0.49	.003
Black	-0.53	0.58	.358	-0.54	0.58	.354	-0.47	0.58	.418	-0.60	0.58	.299	-0.50	0.58	.384
White	0.95	0.63	.136	0.97	0.64	.128	0.99	0.64	.120	0.93	0.63	.142	0.95	0.63	.135
Hispanic	-0.82	0.61	.179	-0.80	0.61	.191	-0.80	0.61	.189	-0.84	0.61	.171	-0.81	0.61	.184
Age at Posttest	0.13	0.03	.000	0.13	0.03	.000	0.13	0.03	.000	0.13	0.03	.000	0.13	0.03	.000
REMA Pretest Score	0.39	0.02	.000	0.39	0.02	.000	0.39	0.02	.000	0.39	0.02	.000	0.39	0.02	.000
Test Lag from School Start	-0.14	0.28	.631	-0.18	0.28	.523	-0.16	0.28	.564	-0.16	0.28	.572	-0.17	0.28	.545
Interval between Test Dates	0.00	0.27	.993	-0.02	0.27	.946	-0.04	0.27	.877	0.00	0.27	.987	-0.02	0.27	.929
Highest Parent Education	0.38	0.13	.004	0.38	0.13	.004	0.39	0.13	.004	0.38	0.13	.004	0.39	0.13	.004
NF Setting Score	0.38	0.23	.095	0.04	0.19	.845	0.40	0.22	.067	0.29	0.21	.159	0.37	0.25	.135
NF Setting Score * Site: Buffalo	-0.42	0.54	.447	-0.34	0.49	.486	-1.18	0.67	.084	-0.04	0.45	.930	-0.46	0.57	.426
NF Setting Score * Site: Boston	-0.37	0.78	.639	-0.60	0.57	.292	-1.23	0.80	.131	-0.40	0.64	.530	-0.16	0.58	.777
NF Setting Score * Site: Nashville	0.00			0.00			0.00			0.00			0.00		

Table A38: Fixed Effects of Near Fidelity Additive Composite on REMA Gain, controlling for COEMET Composite: Time 2

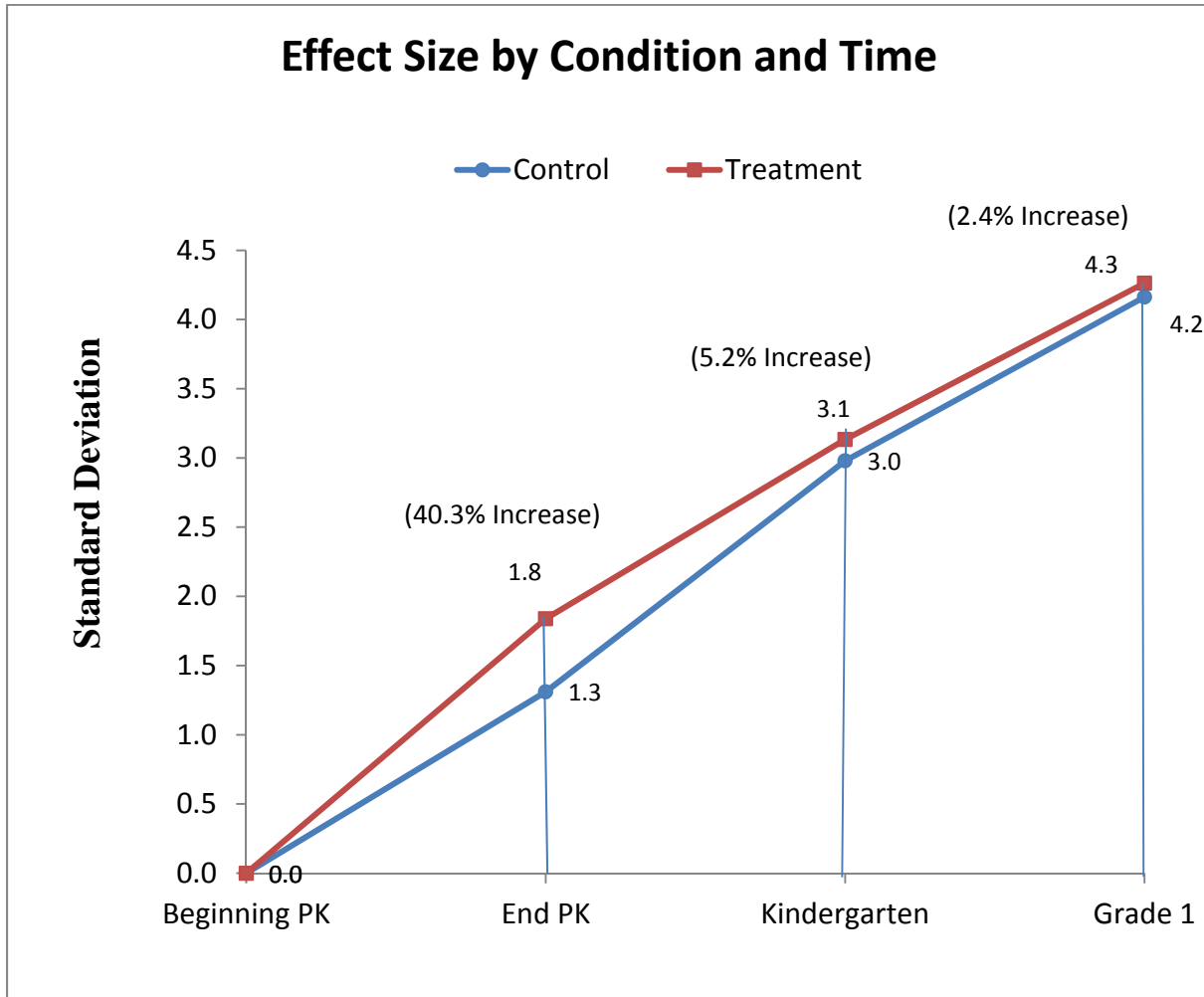
Fixed Effects						
Effect	Full Treatment Sample			Reduced Sample		
	b	SE	p	b	SE	p
Intercept	23.16	2.65	.000	23.39	2.84	.000
Site: Buffalo	1.26	0.46	.009	1.20	0.47	.014
Site: Boston	1.56	0.52	.004	1.68	0.54	.003
Site: Nashville	0.00			0.00		
Male	-0.46	0.22	.037	-0.30	0.23	.196
ELL	1.10	0.44	.013	1.49	0.49	.002
Black	-0.49	0.53	.351	-0.46	0.57	.421
White	0.92	0.58	.110	0.98	0.63	.118
Hispanic	-0.42	0.55	.445	-0.74	0.61	.221
Age at Posttest	0.14	0.03	.000	0.13	0.03	.000
REMA Pretest Score	0.38	0.02	.000	0.39	0.02	.000
Test Lag from School Start	-0.15	0.26	.571	-0.12	0.28	.675
Interval between Test Dates	0.05	0.25	.840	-0.01	0.27	.982
Highest Parent Education	0.48	0.13	.000	0.38	0.13	.004
COEMET Composite	0.91	0.26	.001	0.85	0.27	.002
Near Fidelity Additive Score	0.11	0.09	.246	0.08	0.10	.439

Figure A1: Least Square REMA Means by Condition and Time (Series 1)



Group	Beginning PK	End PK	Kindergarten	Grade 1
Control	38.4	45.2	52.8	59.8
Treatment	38.4	47.9	53.5	60.3

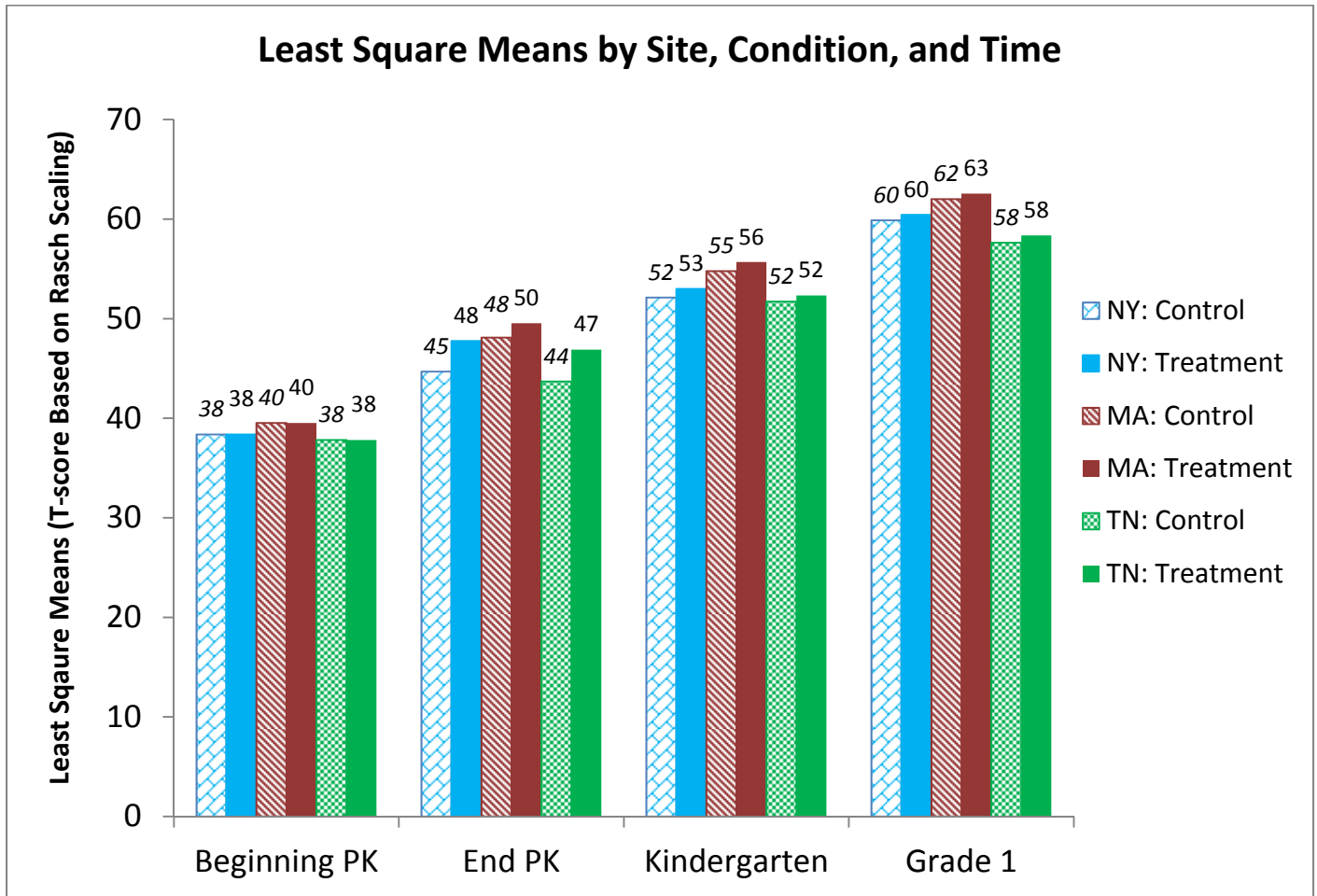
Figure A2: Effect Size on the REMA by Condition and Time (Series 1)



Growth from T1 in T-score				
Control		6.81	14.40	21.43
Treatment		9.56	15.14	21.94

Growth from T1 in ES	Beginning			
	PK	End PK	Kindergarten	Grade 1
Control	0.0	1.3	3.0	4.2
Treatment	0.0	1.8	3.1	4.3
%Increase T vs. C		40.3%	5.2%	2.4%

Figure A3: Least Square REMA Means by Site, Condition, and Time (Series 1)



Site & Condition	Beginning PK	End PK	Kindergarten	Grade 1
NY: Control	38.4	44.7	52.1	59.9
NY: Treatment	38.4	47.8	53.0	60.4
MA: Control	39.5	48.1	54.8	62.0
MA: Treatment	39.5	49.6	55.7	62.6
TN: Control	37.8	43.7	51.7	57.6
TN: Treatment	37.8	46.9	52.3	58.4

Appendix B

Table B1: Sample Sizes across Time (Series 2 – common sample)

Site	N1 (student)	N2 (class)	N3 (school)
Buffalo	530	49	18
Boston	191	21	12
Nashville	384	33	16
Total	1105	103	46

Table B2: Descriptive Statistics by Condition across time: Child Level Variables (Series 2)
(Note: same as Series 1 for the analytic sample for outcome at T3 and T4)

Variable	Control		Treatment	
	Mean	SD	Mean	SD
T-Score - Rasch Converted Pretest (T1)	38.44	5.77	38.39	6.13
T-Score - Rasch Converted Posttest (T2)	44.87	5.35	47.83	4.78
T-Score - Rasch Converted Kindergarten (T3)	52.52	4.81	53.40	4.73
T-Score - Rasch Converted First Grade (T4)	59.60	5.05	60.07	5.11
Race: Black	0.58	0.49	0.63	0.48
Race: White	0.16	0.36	0.19	0.39
Race: Hispanic	0.21	0.40	0.14	0.35
PK ELL Status	0.19	0.39	0.11	0.31
Male	0.48	0.50	0.49	0.50
Parent highest educational level	1.39	0.96	1.51	0.91
Test Lag between school start and pretest in month	1.23	0.47	0.83	0.49
Child Age at REMA T2 (End of PK)	60.16	4.09	60.06	4.10
Child Age at REMA T3 (End of K)	72.15	4.10	71.90	4.11
Child Age at REMA T4 (End of 1st Grade)	83.87	4.02	83.75	4.05
Age interval (month) from REMA T2	-7.15	0.67	-7.71	0.57
Age interval (month) from REMA T3	-19.12	0.59	-19.55	0.60
Age interval (month) from REMA T4	-30.84	0.79	-31.39	0.88
<i>N</i>	521		584	

Table B3: Descriptive Statistics by Site and Condition: Child Level Variables (Series 2)

Variable	Buffalo				Boston				Nashville			
	Control		Treatment		Control		Treatment		Control		Treatment	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
T-Score - Rasch Converted Pretest (T1)	38.64	5.50	38.37	6.30	39.58	6.41	38.81	5.90	37.65	5.75	38.20	6.04
T-Score - Rasch Converted Posttest (T2)	44.78	5.00	47.92	4.57	47.91	4.56	49.60	4.38	43.60	5.62	46.77	4.99
T-Score - Rasch Converted Kindergarten (T3)	52.14	4.48	53.32	4.62	55.41	5.16	55.53	4.15	51.71	4.59	52.39	4.83
T-Score - Rasch Converted First Grade (T4)	59.93	4.56	60.41	4.95	62.79	5.29	62.30	4.44	57.68	4.75	58.42	5.15
Race: Black	0.57	0.50	0.58	0.49	0.31	0.46	0.41	0.49	0.72	0.45	0.82	0.38
Race: White	0.20	0.40	0.31	0.46	0.12	0.32	0.10	0.31	0.12	0.33	0.06	0.23
Race: Hispanic	0.17	0.37	0.08	0.28	0.48	0.50	0.44	0.50	0.13	0.34	0.06	0.24
PK ELL Status	0.15	0.36	0.03	0.16	0.45	0.50	0.41	0.49	0.14	0.34	0.07	0.26
Male	0.50	0.50	0.51	0.50	0.53	0.50	0.50	0.50	0.44	0.50	0.45	0.50
Parent highest educational level	1.53	0.92	1.57	0.92	1.59	1.04	1.44	0.92	1.10	0.92	1.47	0.91
Test Lag between school start and pretest in month	58.78	3.86	58.67	3.70	62.82	3.89	63.07	3.76	60.81	3.71	60.39	3.88
Child Age at REMA T2 (End of PK)	70.69	3.83	70.46	3.70	74.56	3.71	74.91	3.86	73.01	3.86	72.33	3.84
Child Age at REMA T3 (End of K)	82.88	3.86	82.65	3.71	86.42	3.61	86.69	3.82	84.06	3.89	83.73	3.86
Child Age at REMA T4 (End of 1st Grade)	-7.08	0.60	-7.98	0.45	-7.20	1.00	-7.64	0.64	-7.22	0.55	-7.36	0.47
Age interval (month) from REMA T2	-18.99	0.49	-19.76	0.50	-18.94	0.56	-19.48	0.62	-19.37	0.64	-	0.62
Age interval (month) from REMA T3	-31.17	0.61	-31.96	0.55	-30.80	0.57	-31.26	0.56	-30.42	0.89	-	0.83
Age interval (month) from REMA T4	1.38	0.49	0.67	0.48	1.12	0.32	0.84	0.39	1.07	0.44	1.04	0.48
N	251		279		85		106		185		199	

Table B4: Fixed and Random Effects for the REMA Outcome –Common Sample: Time 2 (Series 2)

Fixed Effect

Effect	b	SE	p
Intercept	21.15	3.96	0.0000
Condition_PK	2.96	0.37	0.0000
Site 2: MA	1.92	0.53	0.0006
Site 1: NY	0.85	0.41	0.0425
Site 3: TN (reference)	0.00		
REMA pretest school-mean centered	0.45	0.02	0.0000
REMA pretest school mean	0.36	0.08	0.0001
Age at Time 2	0.17	0.03	0.0000
Age interval between Time 2 and Time 1	0.00	0.23	0.9888
Test Lag from School Start to Pretest	0.01	0.28	0.9610
Ethnicity: Black	-1.50	0.59	0.0106
Ethnicity: White	0.46	0.64	0.4699
Ethnicity: Hispanic	-0.50	0.62	0.4250
English Language Learner Status	0.45	0.46	0.3264
Gender (Male)	-0.42	0.23	0.0689
Parent Education	0.30	0.13	0.0215

Random Effect

Level	Variance	SE	p	ICC
Level 2 (Class)	0.30	0.30	0.1586	0.02
Level 3 (School)	0.58	0.31	0.0306	0.04
Level 1 (Student)	13.93	0.63	0.0000	NA

Random Effect - Unconditional Model

Level	Variance	SE	p	ICC
Level 2 (Class)	0.91	0.58	0.0572	0.03
Level 3 (School)	4.54	1.29	0.0002	0.16
Level 1 (Student)	22.19	0.99	0.0000	NA

Table B5: Fixed and Random Effects for the REMA Outcome -Common Sample: Time 3 (Series 2)

Fixed Effect

Effect	b	SE	p
Intercept	31.89	5.56	0.0000
Condition	0.84	0.27	0.0037
Site 2: MA	1.86	0.41	0.0000
Site 1: NY	0.15	0.29	0.6084
Site 3: TN (reference)	0.00		
REMA pretest school-mean centered	0.38	0.02	0.0000
REMA pretest school mean	0.38	0.06	0.0000
Age at Time 3	0.11	0.03	0.0004
Age interval between Time 3 and Time 1	0.05	0.23	0.8451
Test Lag from School Start to Pretest	-0.24	0.28	0.3909
Ethnicity: Black	-2.68	0.56	0.0000
Ethnicity: White	-0.28	0.61	0.6396
Ethnicity: Hispanic	-1.60	0.60	0.0074
English Language Learner Status	0.62	0.44	0.1581
Gender (Male)	-0.12	0.22	0.6030
Parent Education	0.42	0.13	0.0010

Random Effect

Level	Variance	SE	p	ICC
Level 2 (Class)	0.18	0.26	0.2435	0.01
Level 3 (School)	0.05	0.18	0.3973	0.00
Level 1 (Student)	13.39	0.60	0.0000	NA

Random Effect - Unconditional Model

Level	Variance	SE	p	ICC
Level 2 (Class)	0.65	0.48	0.0890	0.03
Level 3 (School)	2.71	0.89	0.0012	0.12
Level 1 (Student)	19.84	0.89	0.0000	NA

Table B6: Fixed and Random Effects for the REMA Outcome -Common Sample: Time 4 (Series 2)

Fixed Effect

Effect	b	SE	p
Intercept	43.31	6.92	0.0000
Condition	0.49	0.31	0.1269
Site 2: MA	2.93	0.48	0.0000
Site 1: NY	1.43	0.39	0.0005
Site 3: TN (reference)	0.00		
REMA pretest school-mean centered	0.39	0.02	0.0000
REMA pretest school mean	0.42	0.07	0.0000
Age at Time 2	0.04	0.03	0.1924
Age interval between Time 2 and Time 1	0.11	0.20	0.5871
Test Lag from School Start to Pretest	-0.07	0.29	0.8034
Ethnicity: Black	-2.91	0.61	0.0000
Ethnicity: White	-0.52	0.65	0.4287
Ethnicity: Hispanic	-1.60	0.64	0.0132
English Language Learner Status	0.55	0.47	0.2461
Gender (Male)	0.28	0.24	0.2486
Parent Education	0.52	0.14	0.0001

Random Effect

Level	Variance	SE	p	ICC
Level 2 (Class)	0.43	0.36	0.1150	0.03
Level 3 (School)	0.09	0.24	0.3594	0.01
Level 1 (Student)	15.22	0.69	0.0000	NA

Random Effect - Unconditional Model

Level	Variance	SE	p	ICC
Level 2 (Class)	0.82	0.57	0.0759	0.03
Level 3 (School)	3.96	1.25	0.0007	0.15
Level 1 (Student)	21.60	0.97	0.0000	NA

Table B7: Fixed and Random Effects for the REMA Outcome –Common Sample with Condition x Site Interaction: Time 2 (Series 2)

Fixed Effect			
Effect	b	SE	p
Intercept	21.10	3.98	0.0000
Condition	3.21	0.59	0.0000
Site 2: MA	2.57	0.73	0.0008
Site 1: NY	0.85	0.57	0.1471
Site 3: TN (reference)	0.00		
Condition*site 2	-1.27	0.97	0.1980
Condition*site 1	0.04	0.82	0.9636
Condition*site 3	0.00		
REMA pretest school-mean centered	0.45	0.02	0.0000
REMA pretest school mean	0.35	0.08	0.0001
Age at Time 2	0.17	0.03	0.0000
Age interval between Time 2 and Time 1	0.01	0.23	0.9777
Test Lag from School Start to Pretest	0.03	0.29	0.9160
Ethnicity: Black	-1.48	0.59	0.0120
Ethnicity: White	0.48	0.64	0.4486
Ethnicity: Hispanic	-0.48	0.63	0.4388
English Language Learner Status	0.46	0.46	0.3112
Gender (Male)	-0.42	0.23	0.0664
Parent Education	0.30	0.13	0.0247

Random Effect				
Level	Variance	SE	p	ICC
Level 2 (Class)	0.19	0.26	0.2356	0.01
Level 3 (School)	0.05	0.18	0.3938	0.00
Level 1 (Student)	13.39	0.60	0.0000	NA

Type 3 Test		
Effect	FValue	p
Condition	52.29	0.0000
Site	6.19	0.0039
Condition*site	1.11	0.3368
REMA pretest school-mean centered	415.69	0.0000
REMA pretest school mean	18.59	0.0001
Age at Time 2	27.92	0.0000
Age interval between Time 2 and Time 1	0.00	0.9777
Test Lag from School Start to Pretest	0.01	0.9160
Ethnicity: Black	6.33	0.0120
Ethnicity: White	0.57	0.4486
Ethnicity: Hispanic	0.60	0.4388
English Language Learner Status	1.03	0.3112
Gender (Male)	3.38	0.0664
Parent Education	5.06	0.0247

Table B8: Fixed and Random Effects for the REMA Outcome -Common Sample with Condition x Site Interaction: Time 3 (Series 2)

Fixed Effect			
Effect	b	SE	p
Intercept	32.58	5.63	0.0000
Condition	0.56	0.43	0.2003
Site 2: MA	1.85	0.58	0.0020
Site 1: NY	-0.18	0.42	0.6648
Site 3: TN (reference)	0.00		
Condition*site 2	0.02	0.75	0.9766
Condition*site 1	0.67	0.60	0.2695
Condition*site 3	0.00		
REMA pretest school-mean centered	0.38	0.02	0.0000
REMA pretest school mean	0.39	0.06	0.0000
Age at Time 3	0.11	0.03	0.0003
Age interval between Time 3 and Time 1	0.08	0.24	0.7256
Test Lag from School Start to Pretest	-0.18	0.28	0.5366
Ethnicity: Black	-2.69	0.57	0.0000
Ethnicity: White	-0.34	0.61	0.5744
Ethnicity: Hispanic	-1.62	0.60	0.0067
English Language Learner Status	0.63	0.44	0.1477
Gender (Male)	-0.12	0.22	0.6067
Parent Education	0.42	0.13	0.0009

Random Effect				
Level	Variance	SE	p	ICC
Level 2 (Class)	0.19	0.26	0.2356	0.01
Level 3 (School)	0.05	0.18	0.3938	0.00
Level 1 (Student)	13.39	0.60	0.0000	NA

Type 3 Test		
Effect	FValue	p
Condition	7.41	0.0093
Site	7.28	0.0016
Condition*site	0.79	0.4627
REMA pretest school-mean centered	317.90	0.0000
REMA pretest school mean	39.09	0.0000
Age at Time 3	12.92	0.0003
Age interval between Time 3 and Time 1	0.12	0.7256
Test Lag from School Start to Pretest	0.38	0.5366
Ethnicity: Black	22.68	0.0000
Ethnicity: White	0.32	0.5744
Ethnicity: Hispanic	7.38	0.0067
English Language Learner Status	2.10	0.1477
Gender (Male)	0.27	0.6067
Parent Education	11.08	0.0009

Table B9: Fixed and Random Effects for the REMA Outcome -Common Sample with Condition x Site Interaction: Time 4 (Series 2)

Fixed Effect			
Effect	b	SE	p
Intercept	43.45	6.97	0.0000
Condition	0.56	0.51	0.2788
Site 2: MA	3.20	0.67	0.0000
Site 1: NY	1.41	0.53	0.0099
Site 3: TN (reference)	0.00		
Condition *site 2	-0.50	0.87	0.5656
Condition *site 1	0.06	0.70	0.9365
Condition *site 3	0.00		
REMA pretest school-mean centered	0.39	0.02	0.0000
REMA pretest school mean	0.42	0.07	0.0000
Age at Time 4	0.04	0.03	0.1881
Age interval between Time 4 and Time 1	0.11	0.20	0.5702
Test Lag from School Start to Pretest	-0.06	0.29	0.8489
Ethnicity: Black	-2.90	0.61	0.0000
Ethnicity: White	-0.50	0.66	0.4461
Ethnicity: Hispanic	-1.59	0.65	0.0137
English Language Learner Status	0.55	0.47	0.2442
Gender (Male)	0.28	0.24	0.2481
Parent Education	0.52	0.14	0.0002

Random Effect				
Level	Variance	SE	p	ICC
Level 2 (Class)	0.42	0.36	0.1164	0.03
Level 3 (School)	0.14	0.26	0.2931	0.01
Level 1 (Student)	15.21	0.68	0.0000	NA

Type 3 Test		
Effect	FValue	p
Condition	7.41	0.0093
site	7.28	0.0016
Condition *site	0.79	0.4627
REMA pretest school-mean centered	317.90	0.0000
REMA pretest school mean	39.09	0.0000
Age at Time 4	12.92	0.0003
Age interval between Time 4 and Time 1	0.12	0.7256
Test Lag from School Start to Pretest	0.38	0.5366
Ethnicity: Black	22.68	0.0000
Ethnicity: White	0.32	0.5744
Ethnicity: Hispanic	7.38	0.0067
English Language Learner Status	2.10	0.1477
Gender (Male)	0.27	0.6067
Parent Education	11.08	0.0009

Table B10: Fixed and Random Effects for the REMA Outcome -By site: Time 2 (Series 2)

Fixed Effect									
Effect	Site 1: NY			Site 2: MA			Site 3: TN		
	b	SE	p	b	SE	p	b	SE	p
Intercept	16.03	9.06	0.0910	22.16	6.91	0.0023	33.43	6.43	0.0000
Condition	3.09	0.75	0.0006	2.20	0.66	0.0216	3.23	0.48	0.0000
REMA pretest school-mean centered	0.36	0.03	0.0000	0.38	0.06	0.0000	0.61	0.04	0.0000
REMA pretest school mean	0.40	0.21	0.0756	0.39	0.10	0.0064	0.21	0.12	0.1099
Age at Time 2	0.21	0.04	0.0000	0.16	0.07	0.0363	0.11	0.06	0.0575
Age interval between Time 2 and Time 1	-0.12	0.35	0.7385	0.04	0.37	0.9228	0.38	0.53	0.4716
Test Lag from School Start to Pretest	0.21	0.37	0.5700	0.19	0.81	0.8160	-0.38	0.57	0.5071
Ethnicity: Black	-0.93	0.86	0.2798	-1.22	1.23	0.3218	-1.23	1.18	0.3004
Ethnicity: White	1.61	0.91	0.0779	-0.61	1.39	0.6602	-0.25	1.35	0.8507
Ethnicity: Hispanic	-0.12	0.94	0.9021	-0.41	1.21	0.7324	-0.86	1.62	0.5963
English Language Learner Status	-0.38	0.76	0.6200	1.06	0.69	0.1246	1.11	1.70	0.5144
Gender (Male)	-0.45	0.32	0.1581	-0.12	0.55	0.8256	-0.46	0.41	0.2613
Parent Education	0.29	0.19	0.1153	0.58	0.31	0.0652	0.12	0.24	0.6129

Random Effect									
Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	0.88	0.53	0.0498	0.00			0.00		
Level 3 (School)	1.16	0.77	0.0648	0.09	0.66	0.4480	0.19	0.32	0.2741
Level 1 (Student)	12.65	0.82	0.0000	13.39	1.48	0.0000	14.99	1.12	0.0000

Random Effect - Unconditional Model									
Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	1.08	0.76	0.0766	0.00			1.10	1.24	0.188026
Level 3 (School)	4.40	1.85	0.008642	2.48	1.73	0.075366	1.94	1.42	0.086581
Level 1 (Student)	19.46	1.26	1.91E-54	18.54	1.97	2.07E-21	27.68	2.10	3.99E-40

Table B11: Fixed and Random Effects for the REMA Outcome -By site: Time 3 (Series 2)

Fixed Effect									
Effect	Site 1: NY			Site 2: MA			Site 3: TN		
	b	SE	p	b	SE	p	b	SE	p
Intercept	26.84	9.57	0.0059	53.91	12.25	0.0000	33.67	9.44	0.0005
Condition	1.02	0.52	0.0610	0.77	0.85	0.3912	0.73	0.40	0.0779
REMA pretest school-mean centered	0.32	0.03	0.0000	0.31	0.06	0.0000	0.50	0.04	0.0000
REMA pretest school mean	0.34	0.14	0.0252	0.38	0.13	0.0168	0.35	0.11	0.0017
Age at Time 3	0.18	0.04	0.0001	0.01	0.07	0.8449	0.04	0.05	0.3894
Age interval between Time 3 and Time 1	-0.04	0.37	0.9053	0.72	0.53	0.1727	-0.13	0.39	0.7414
Test Lag from School Start to Pretest	-0.31	0.37	0.4052	-0.44	0.82	0.5931	0.29	0.53	0.5784
Ethnicity: Black	-2.66	0.84	0.0016	-3.64	1.22	0.0032	-1.46	1.12	0.1939
Ethnicity: White	0.16	0.89	0.8535	-2.24	1.38	0.1053	0.85	1.26	0.5005
Ethnicity: Hispanic	-1.08	0.90	0.2303	-2.57	1.20	0.0334	-1.73	1.53	0.2585
English Language Learner Status	0.56	0.72	0.4348	1.28	0.69	0.0631	0.69	1.61	0.6672
Gender (Male)	0.12	0.32	0.7078	-0.58	0.54	0.2909	-0.10	0.39	0.7993
Parent Education	0.51	0.18	0.0059	1.29	0.30	0.0000	-0.18	0.22	0.4247

Random Effect									
Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	0.49	0.44	0.1369	0.00			0.05	0.29	0.4306
Level 3 (School)	0.13	0.31	0.3338	0.94	1.03	0.1800	0.00		
Level 1 (Student)	12.57	0.83	0.0000	13.08	1.43	0.0000	13.36	1.01	0.0000

Random Effect - Unconditional Model									
Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	0.84	0.67	0.1061	0.00			0.43	0.56	0.2196
Level 3 (School)	1.85	0.95	0.0263	2.77	1.86	0.0686	0.00		
Level 1 (Student)	18.40	1.19	0.0000	19.18	2.03	0.0000	21.88	1.64	0.0000

Table B12: Fixed and Random Effects for the REMA Outcome -By site: Time 4 (Series 2)

Fixed Effect

Effect	Site 1: NY			Site 2: MA			Site 3: TN		
	b	SE	p	b	SE	p	b	SE	p
Intercept	49.73	12.41	0.0001	88.84	20.22	0.0000	45.73	10.16	0.0000
Condition	0.55	0.60	0.3686	0.55	1.00	0.5958	0.73	0.43	0.0898
REMA pretest school-mean centered	0.33	0.03	0.0000	0.33	0.06	0.0000	0.51	0.04	0.0000
REMA pretest school mean	0.32	0.17	0.0750	0.58	0.15	0.0044	0.26	0.11	0.0203
Age at Time 4	0.09	0.05	0.0526	-0.16	0.08	0.0358	0.04	0.06	0.4734
Age interval between Time 4 and Time 1	0.27	0.35	0.4390	1.17	0.58	0.0436	0.00	0.27	0.9884
Test Lag from School Start to Pretest	-0.16	0.40	0.6923	-0.40	0.86	0.6405	0.21	0.49	0.6720
Ethnicity: Black	-3.11	0.91	0.0007	-3.24	1.28	0.0126	-1.77	1.20	0.1424
Ethnicity: White	-0.26	0.97	0.7903	-0.94	1.43	0.5102	0.10	1.36	0.9429
Ethnicity: Hispanic	-1.61	0.99	0.1047	-0.91	1.26	0.4720	-1.68	1.65	0.3098
English Language Learner Status	0.20	0.78	0.7990	1.33	0.71	0.0619	0.28	1.74	0.8705
Gender (Male)	0.40	0.34	0.2442	0.51	0.56	0.3724	0.23	0.42	0.5806
Parent Education	0.66	0.20	0.0009	1.37	0.31	0.0000	-0.10	0.24	0.6641

Random Effect

Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	0.83	0.56	0.0686	0.45	1.28	0.3632	0.00		
Level 3 (School)	0.32	0.47	0.2513	1.18	1.73	0.2469	0.00		
Level 1 (Student)	14.32	0.93	0.0000	13.37	1.51	0.0000	15.86	1.16	0.0000

Random Effect - Unconditional Model

Level	Site 1: NY			Site 2: MA			Site 3: TN		
	Variance	SE	p	Variance	SE	p	Variance	SE	p
Level 2 (Class)	0.91	0.72	0.1019	1.23	1.98	0.2680	0.00		
Level 3 (School)	1.70	0.95	0.0364	4.12	3.33	0.1081	0.00		
Level 1 (Student)	20.09	1.29	0.0000	19.16	2.08	0.0000	24.68	1.78	0.0000

Table B13: Summary of Moderator Analyses for REMA Outcome: Time 2 (Series 2)

Moderator	Male		ELL		Pretest		Age_T2		Black		Hispanic		White	
	b	p	b	p	b	p	b	p	b	p	b	p	b	p
Intercept	44.78	0.0000	44.78	0.0000	44.89	0.0000	45.00	0.0000	45.19	0.0000	44.62	0.0000	44.75	0.0000
Condition	2.74	0.0000	3.02	0.0000	2.99	0.0000	3.02	0.0000	2.13	0.0000	3.20	0.0000	3.08	0.0000
Condition*Moderator	0.48	0.3005	-0.34	0.6418	-0.10	0.0090	-0.13	0.0255	1.38	0.0111	-1.29	0.0613	-0.69	0.3197
Site 2: MA	1.76	0.0010	1.77	0.0009	1.71	0.0012	1.78	0.0008	1.80	0.0005	1.80	0.0006	1.77	0.0009
Site 1: NY	0.80	0.0544	0.80	0.0529	0.79	0.0553	0.84	0.0424	0.82	0.0401	0.80	0.0477	0.83	0.0446
Site 3: TN	0		0		0		0		0		0		0	
REMA pretest grand-mean centered	0.44	0.0000	0.44	0.0000	0.50	0.0000	0.44	0.0000	0.44	0.0000	0.44	0.0000	0.44	0.0000
Age at Time 2	0.17	0.0000	0.17	0.0000	0.17	0.0000	0.24	0.0000	0.17	0.0000	0.17	0.0000	0.17	0.0000
Age interval between Time 2 and Time 1	-0.02	0.9286	0.00	0.9864	0.01	0.9665	0.03	0.8882	-0.01	0.9599	-0.01	0.9501	-0.01	0.9782
Test Lag from School Start to Pretest	0.02	0.9429	-0.01	0.9715	0.01	0.9748	0.02	0.9383	-0.05	0.8619	-0.02	0.9393	-0.02	0.9541
Ethnicity: Black	-1.49	0.0112	-1.50	0.0109	-1.48	0.0114	-1.50	0.0104	-2.17	0.0007	-1.50	0.0106	-1.50	0.0108
Ethnicity: White	0.44	0.4859	0.43	0.5026	0.47	0.4602	0.39	0.5359	0.45	0.4750	0.42	0.5096	0.77	0.2871
Ethnicity: Hispanic	-0.44	0.4773	-0.44	0.4830	-0.41	0.5050	-0.46	0.4642	-0.44	0.4805	0.16	0.8221	-0.44	0.4845
English Language Learner Status	0.48	0.2885	0.61	0.2638	0.54	0.2352	0.47	0.3012	0.44	0.3316	0.45	0.3255	0.48	0.2904
Gender (Male)	-0.68	0.0430	-0.43	0.0645	-0.45	0.0530	-0.41	0.0737	-0.44	0.0577	-0.42	0.0681	-0.44	0.0593
Parent Education	0.29	0.0256	0.30	0.0247	0.28	0.0306	0.30	0.0234	0.28	0.0303	0.29	0.0257	0.29	0.0280

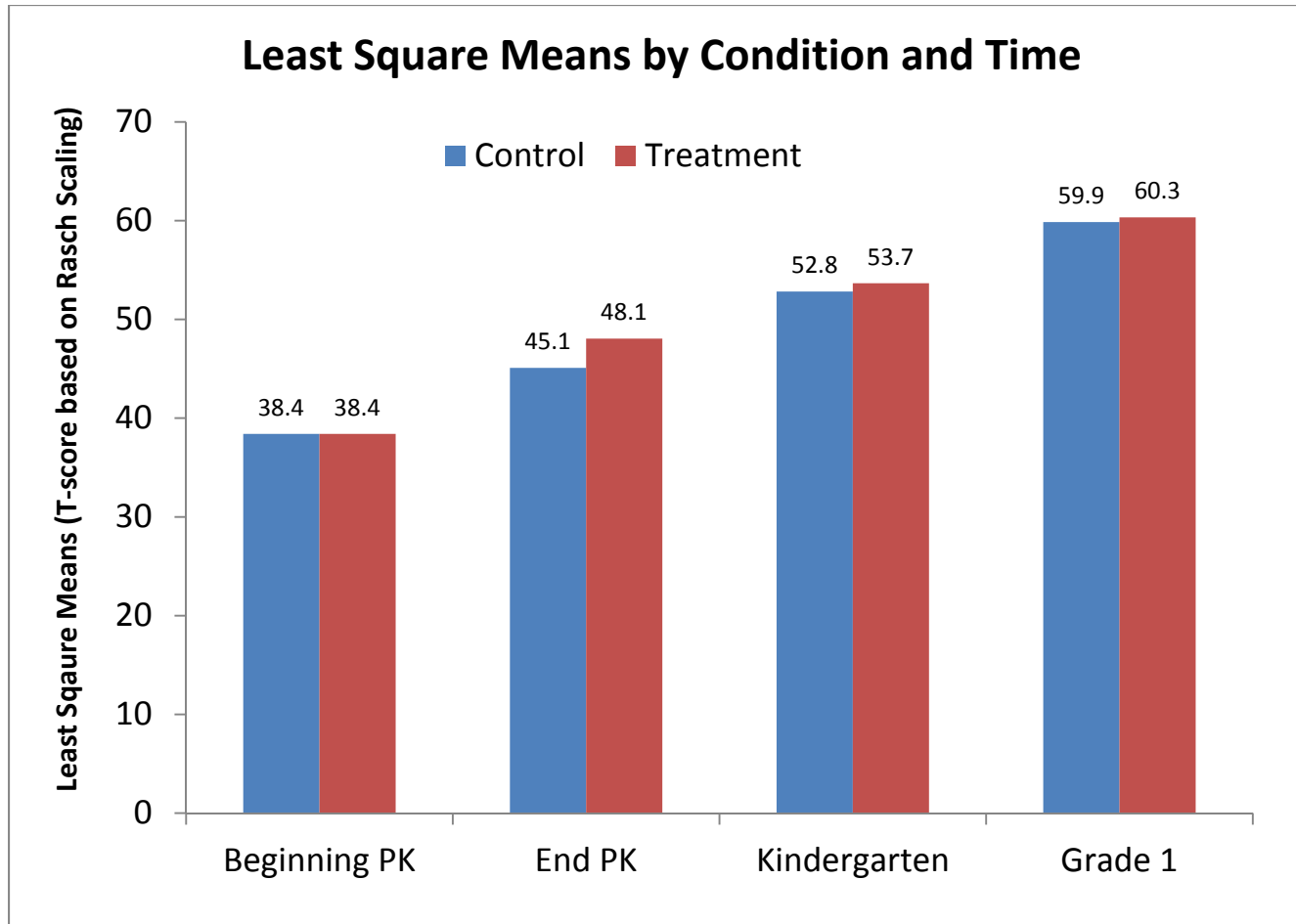
Table B14: Summary of Moderator Analyses for REMA Outcome: Time 3 (Series 2)

Moderator	Male		ELL		Pretest		Age_T3		Black		Hispanic		White	
	b	p	b	p	b	p	b	p	b	p	b	p	b	p
Intercept	54.59	0.0000	54.43	0.0000	54.84	0.0000	54.56	0.0000	54.64	0.0000	54.50	0.0000	54.57	0.0000
Condition	0.94	0.0072	0.91	0.0026	0.85	0.0030	0.86	0.0025	0.35	0.3932	0.90	0.0031	0.85	0.0048
Condition*Moderator	-0.21	0.6375	-0.52	0.4386	-0.04	0.2809	-0.08	0.1368	0.78	0.1122	-0.40	0.5258	-0.08	0.8948
Site 2: MA	1.86	0.0000	1.89	0.0000	1.84	0.0000	1.90	0.0000	1.90	0.0000	1.89	0.0000	1.87	0.0000
Site 1: NY	0.16	0.5952	0.14	0.6245	0.14	0.6227	0.17	0.5608	0.16	0.5737	0.15	0.6039	0.16	0.5951
Site 3: TN	0		0		0		0		0		0		0	
REMA pretest grand-mean centered	0.38	0.0000	0.38	0.0000	0.41	0.0000	0.38	0.0000	0.38	0.0000	0.38	0.0000	0.38	0.0000
Age at Time 3	0.11	0.0003	0.11	0.0003	0.11	0.0004	0.15	0.0003	0.11	0.0004	0.11	0.0003	0.11	0.0004
Age interval between Time 3 and Time 1	0.05	0.8363	0.04	0.8623	0.06	0.7983	0.05	0.8338	0.03	0.8808	0.04	0.8536	0.05	0.8450
Test Lag from School Start to Pretest	-0.25	0.3711	-0.24	0.3868	-0.24	0.3888	-0.20	0.4754	-0.27	0.3324	-0.25	0.3809	-0.24	0.3844
Ethnicity: Black	-2.68	0.0000	-2.67	0.0000	-2.66	0.0000	-2.68	0.0000	-3.06	0.0000	-2.68	0.0000	-2.68	0.0000
Ethnicity: White	-0.30	0.6252	-0.28	0.6468	-0.26	0.6698	-0.31	0.6074	-0.25	0.6801	-0.29	0.6295	-0.24	0.7277
Ethnicity: Hispanic	-1.60	0.0072	-1.59	0.0076	-1.58	0.0078	-1.61	0.0067	-1.61	0.0069	-1.43	0.0299	-1.60	0.0072
English Language Learner Status	0.61	0.1602	0.83	0.1067	0.65	0.1370	0.63	0.1479	0.59	0.1786	0.61	0.1636	0.62	0.1553
Gender (Male)	-0.01	0.9839	-0.12	0.5951	-0.13	0.5774	-0.11	0.6325	-0.13	0.5770	-0.11	0.6093	-0.12	0.5997
Parent Education	0.42	0.0010	0.42	0.0009	0.41	0.0012	0.42	0.0009	0.41	0.0013	0.42	0.0010	0.42	0.0010

Table B15: Summary of Moderator Analyses for REMA Outcome: Time 4 (Series 2)

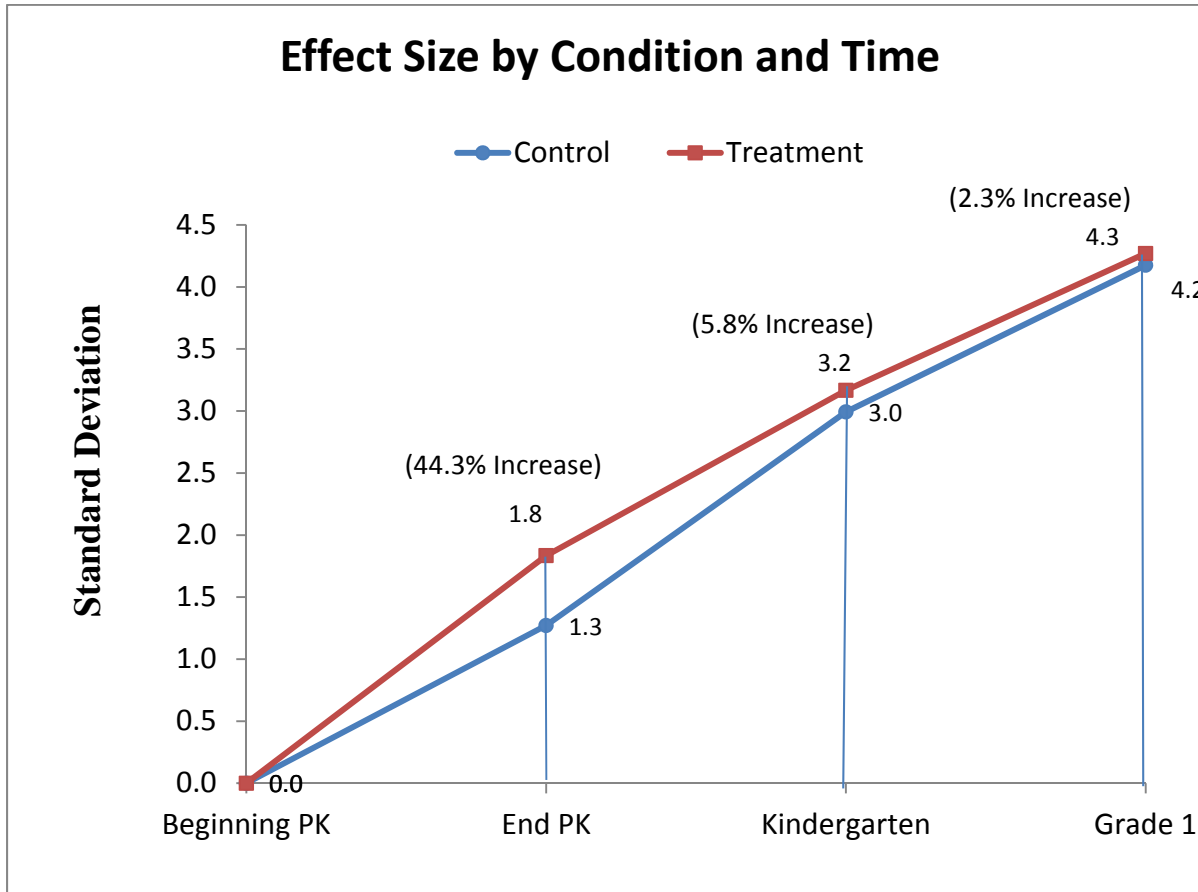
Moderator	Male		ELL		Pretest		Age_T4		Black		Hispanic		White	
	b	p	b	p	b	p	b	p	b	p	b	p	b	p
Intercept	62.85	0.0000	62.74	0.0000	62.86	0.0000	62.85	0.0000	62.61	0.0000	62.78	0.0000	62.91	0.0000
Condition	0.59	0.1284	0.53	0.1162	0.49	0.1250	0.49	0.1251	0.07	0.8791	0.55	0.1018	0.47	0.1620
Condition*Moderator	-0.22	0.6501	-0.27	0.7112	0.00	0.9395	0.00	0.9680	0.66	0.2173	-0.38	0.5764	0.11	0.8716
Site 2: MA	2.98	0.0000	2.99	0.0000	2.98	0.0000	2.98	0.0000	3.00	0.0000	3.00	0.0000	2.98	0.0000
Site 1: NY	1.45	0.0004	1.43	0.0004	1.44	0.0004	1.44	0.0004	1.44	0.0004	1.44	0.0004	1.44	0.0004
Site 3: TN	0		0		0		0		0		0		0	
REMA pretest grand-mean centered	0.39	0.0000	0.39	0.0000	0.39	0.0000	0.39	0.0000	0.39	0.0000	0.39	0.0000	0.39	0.0000
Age at Time 4	0.04	0.1904	0.04	0.1953	0.04	0.1964	0.04	0.3400	0.04	0.2040	0.04	0.1941	0.04	0.1944
Age interval between Time 4 and Time 1	0.11	0.5917	0.10	0.6059	0.11	0.5949	0.10	0.5960	0.09	0.6517	0.10	0.6016	0.11	0.5902
Test Lag from School Start to Pretest	-0.07	0.8106	-0.06	0.8317	-0.06	0.8353	-0.06	0.8364	-0.08	0.7735	-0.06	0.8199	-0.06	0.8404
Ethnicity: Black	-2.92	0.0000	-2.91	0.0000	-2.91	0.0000	-2.91	0.0000	-3.24	0.0000	-2.91	0.0000	-2.91	0.0000
Ethnicity: White	-0.51	0.4368	-0.49	0.4491	-0.50	0.4464	-0.50	0.4437	-0.47	0.4668	-0.50	0.4402	-0.56	0.4538
Ethnicity: Hispanic	-1.61	0.0120	-1.61	0.0123	-1.61	0.0120	-1.61	0.0119	-1.62	0.0117	-1.44	0.0434	-1.61	0.0118
English Language Learner Status	0.53	0.2579	0.65	0.2438	0.54	0.2506	0.54	0.2514	0.52	0.2713	0.53	0.2600	0.54	0.2539
Gender (Male)	0.39	0.2577	0.28	0.2479	0.28	0.2468	0.28	0.2452	0.27	0.2594	0.28	0.2431	0.28	0.2424
Parent Education	0.53	0.0001	0.53	0.0001	0.53	0.0001	0.53	0.0001	0.52	0.0001	0.53	0.0001	0.53	0.0001

Figure B1: Least Square REMA Means by Condition and Time (Series 2)



Group	Beginning PK	End PK	Kindergarten	Grade 1
Control	38.4	45.1	52.8	59.9
Treatment	38.4	48.1	53.7	60.3

Figure B2: Effect Size on the REMA by Condition and Time (Series 2)



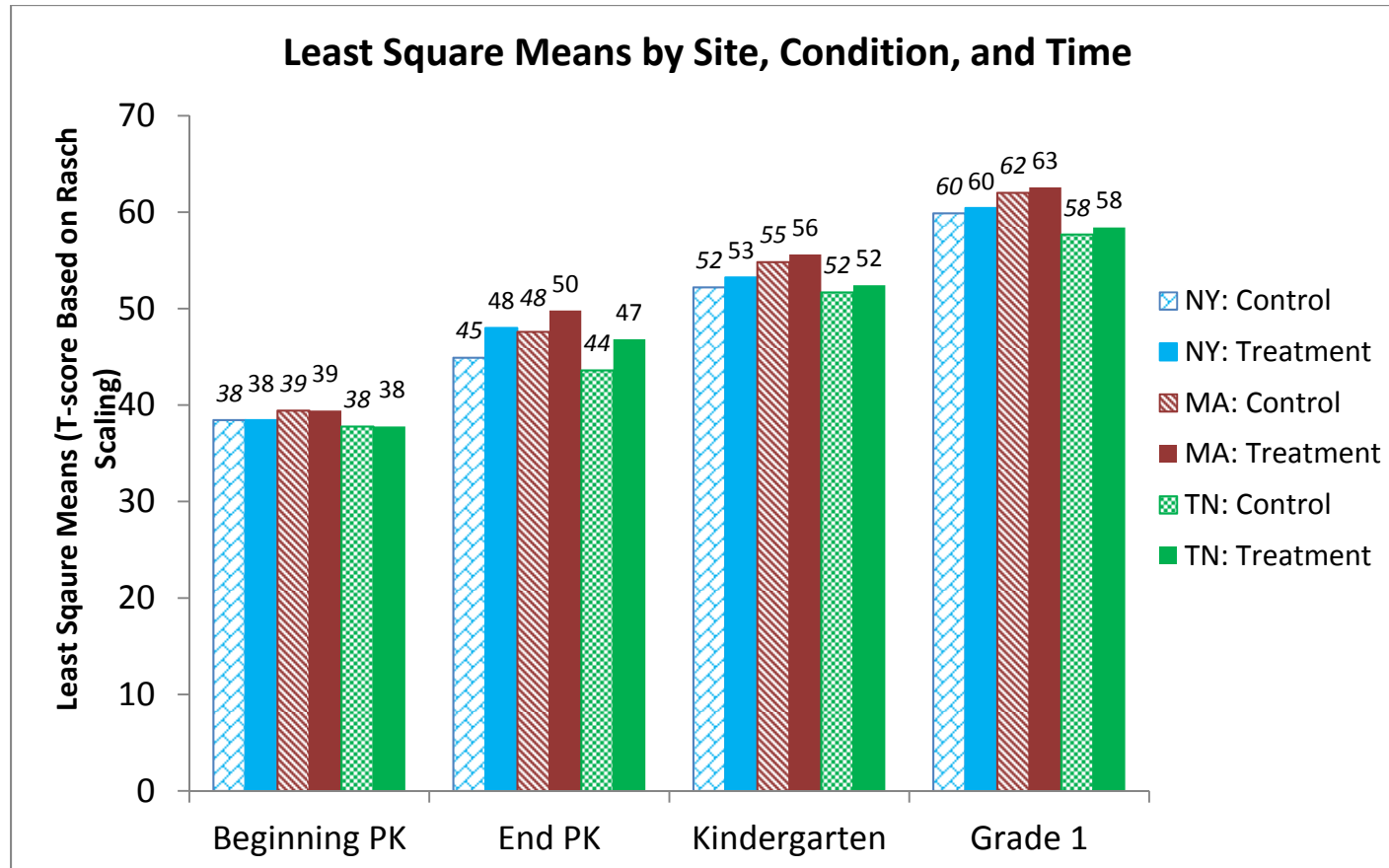
Growth from T1 in T-score

Control	6.69	14.41	21.45
Treatment	9.64	15.25	21.93

Growth from T1 in ES

	Beginning PK	End PK	Kindergarten	Grade 1
Control	0.0	1.3	3.0	4.2
Treatment	0.0	1.8	3.2	4.3
%Increase T vs. C		44.3%	5.8%	2.3%

Figure B3: Least Square REMA Means by Site, Condition, and Time (Series 2)



Site & Condition	Beginning PK	End PK	Kindergarten	Grade 1
NY: Control	38.4	44.9	52.2	59.9
NY: Treatment	38.4	48.0	53.2	60.4
MA: Control	39.4	47.6	54.8	62.0
MA: Treatment	39.4	49.8	55.6	62.6
TN: Control	37.8	43.6	51.7	57.7
TN: Treatment	37.8	46.8	52.4	58.4