Corporation for
NATIONAL & COMMUNITY SERVICE

OUTCOME EVALUATION

APPENDICES

Minnesota Reading Corps

PreK Program

MARCH 2015

Outcome Evaluation of the Minnesota Reading Corps PreK Program

Appendices of Final Report

February 2015

Authors

This study was conducted by researchers from NORC at the University of Chicago and TIES:

Carrie E. Markovitz, Ph.D., Principal Research Scientist, NORC at the University of Chicago
Marc W. Hernandez, Ph.D., Senior Research Scientist, NORC at the University of Chicago
Eric C. Hedberg, Ph.D., Senior Research Scientist, NORC at the University of Chicago
Benjamin Silberglitt, Ph.D., Director of Software Applications, TIES

This report represents the work and perspectives of the authors and is the product of professional research. It does not represent the position or opinions of CNCS, the federal government, or the reviewers.

Citation

Markovitz, C.; Hernandez, M.; Hedberg, E.; Silberglitt, B. (2015). *Outcome Evaluation of the Minnesota Reading Corps PreK Program: Appendices*. NORC at the University of Chicago: Chicago, IL.

List of Appendices

Appendix A.1: Logic Model for the Process and Impact Evaluation of MRC 1

Appendix A.2: School Matching Validity Analysis 2

Appendix A.3: Detailed Methodology 11

Appendix B.1: Fall-Winter Models and Effects Tables 16

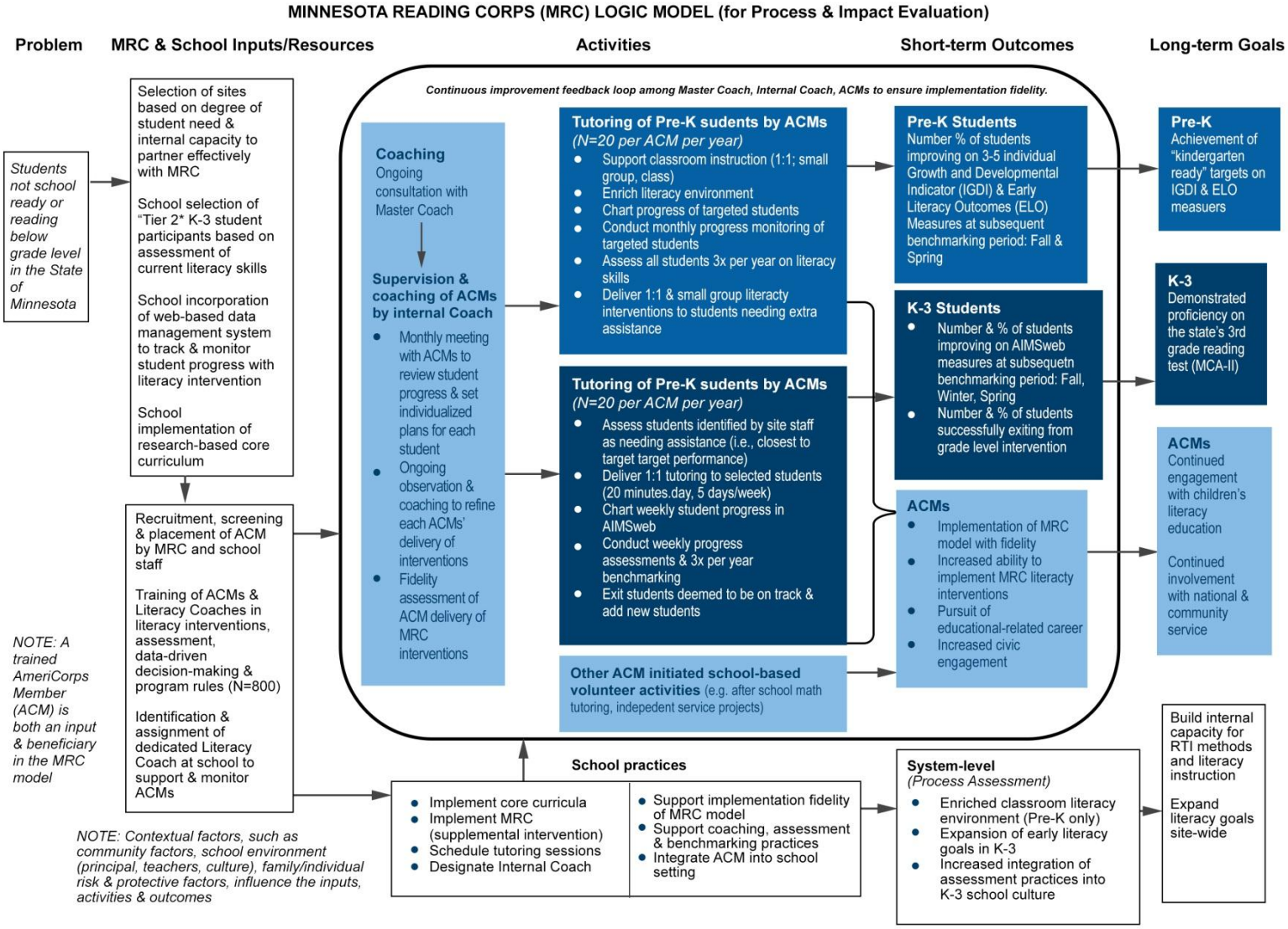
Appendix B.2: Fall-Spring Models and Effects Tables 25

Appendix C: Findings from Robustness Analysis 34

Appendix D: IES What Works Clearinghouse Analysis 36

Appendix E: Glossary 37

Appendix A.1: Logic Model for the Process and Impact Evaluation of MRC



Appendix A.2: School Matching Validity Analysis

As discussed in the report, site pairs were formed by matching each comparison site to a Minnesota Reading Corps PreK program site (intervention) on a select group of educationally important pre-intervention characteristics[1] (i.e., urbanicity, institution type, classroom composition of student ages, percentage of students eligible for FRPL[2], and percentage of students who were DLLs). The analyses below provide confirmation of the validity of our matching process.

Site Comparison by Study Participation

Although 39 Minnesota Reading Corps program sites were sampled to participate in the PreK outcome evaluation, only 25 comparison sites were identified, which matched those of the sampled program sites on a prescribed set of educationally significant baseline characteristics (e.g., urbanicity, program type, student age mix, etc.). Thus, 25 pairs of sites (i.e., 25 Minnesota Reading Corps program sites and 25 comparison sites) participated in the evaluation during the 2013-2014 school year. In order to establish that the 25 selected program sites did not substantively differ from the 14 unselected program sites, we performed bivariate logistic regressions predicting selection by the following characteristics (*p*-values of model chi-square statistics in parentheses):

- Average hours of class time per week ($p = 0.9603$)
- Teacher/student ratio ($p = 0.5538$)
- Percentage free and reduced price lunch ($p = 0.0636$)
- Program type ($p = 0.7726$)
- Age of students ($p = 0.5028$)
- Average hours of class time per day ($p = 0.7680$)

Based on the *p*-values provided above (all of which are >.05), we are confident that our 25 program sites are representative of the pool of 39 potential candidates. With this result, we believe our results achieve external validity.

[1] These variables were based on WWC recommendations, administrative data collected by Minnesota schools at the time, the literature on predictors of preschool emergent-literacy outcomes, and consultation with the project's Technical Working Group (TWG).

[2] A measure of household poverty and socio-economic status.

Differences in Fall Outcomes by Site Pair

The purpose of this analysis is to identify pairs in which the difference between the averages of the intervention and comparison sites was statistically significant on Fall (baseline) outcomes. This was accomplished using a simple regression model for each site pair, whereby the Fall outcome is a function of an intervention indicator (MRC_j) and age is a covariate (Age_{ij}).

$$y_{ij} = \beta_0 + \beta_1 MRC_j + \beta_2 Age_{ij} + e_{ij}$$

The coefficient of the intervention indicator (β_1) measures the difference in mean outcomes in Fall, and the confidence interval is constructed with the value of t appropriate for 2 degrees of freedom and an α of 0.05.

$$CI_{1-\alpha} = \beta_1 \pm t_{\frac{\alpha}{2},2} \times SE(\beta_1)$$

If the confidence interval encompasses zero, then there is no statistical difference between the intervention site and the comparison site. However, if the confidence interval falls outside of 0, then it is highly likely that there is a difference between sites.

The charts below are organized as follows. Each pair, identified with the pair number on the x-axis, is assigned a vertical red line. This line is the confidence interval of the difference. Balance between the intervention and comparison site is indicated when this red line passes zero on the y-axis. Thus, the vast majority of site pairs on most outcomes are balanced for 3-year old students, while some site pairs are imbalanced on certain outcomes for 4- and 5-year old students.

Figure 1: Balance Tests for 3-year-olds

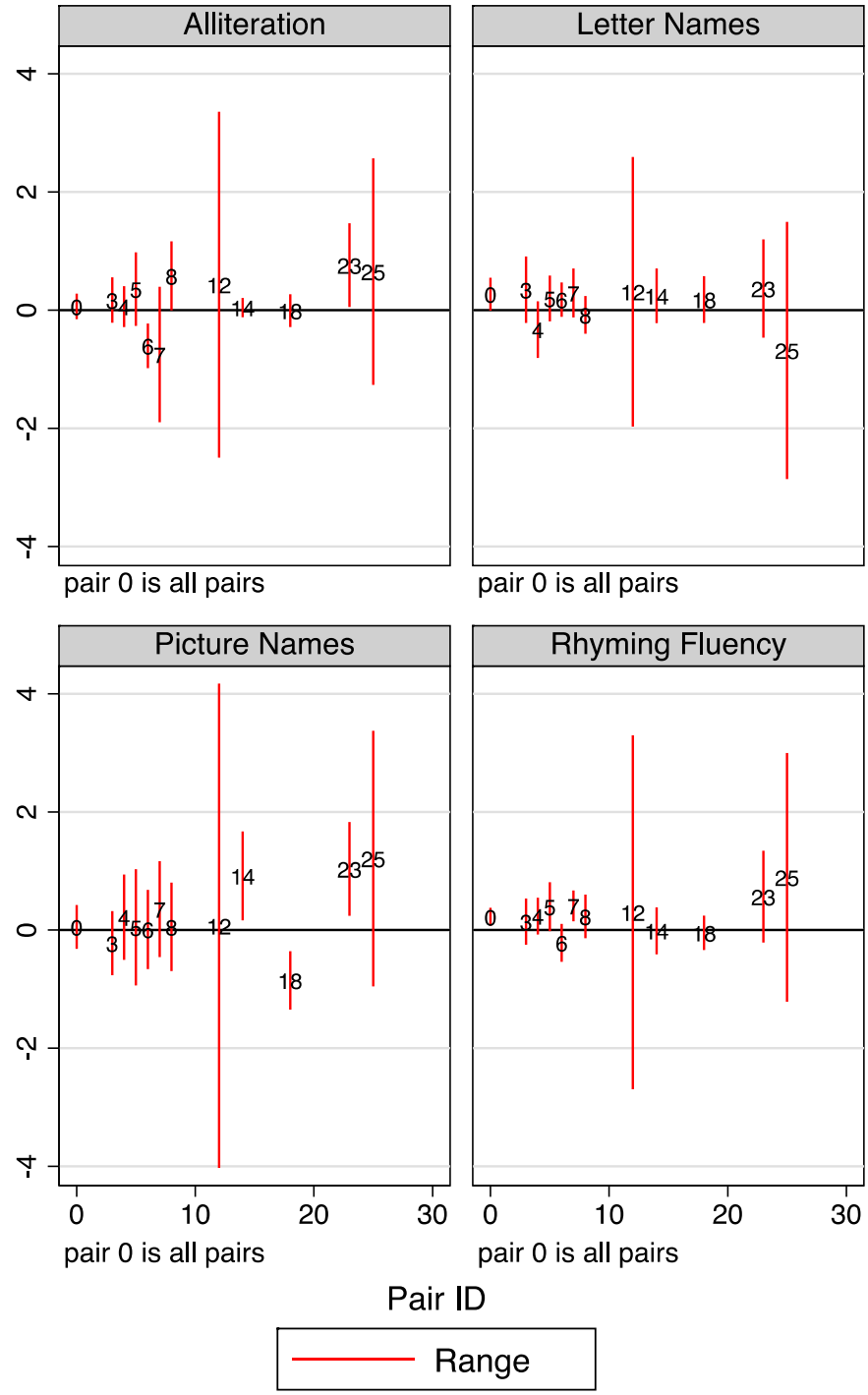
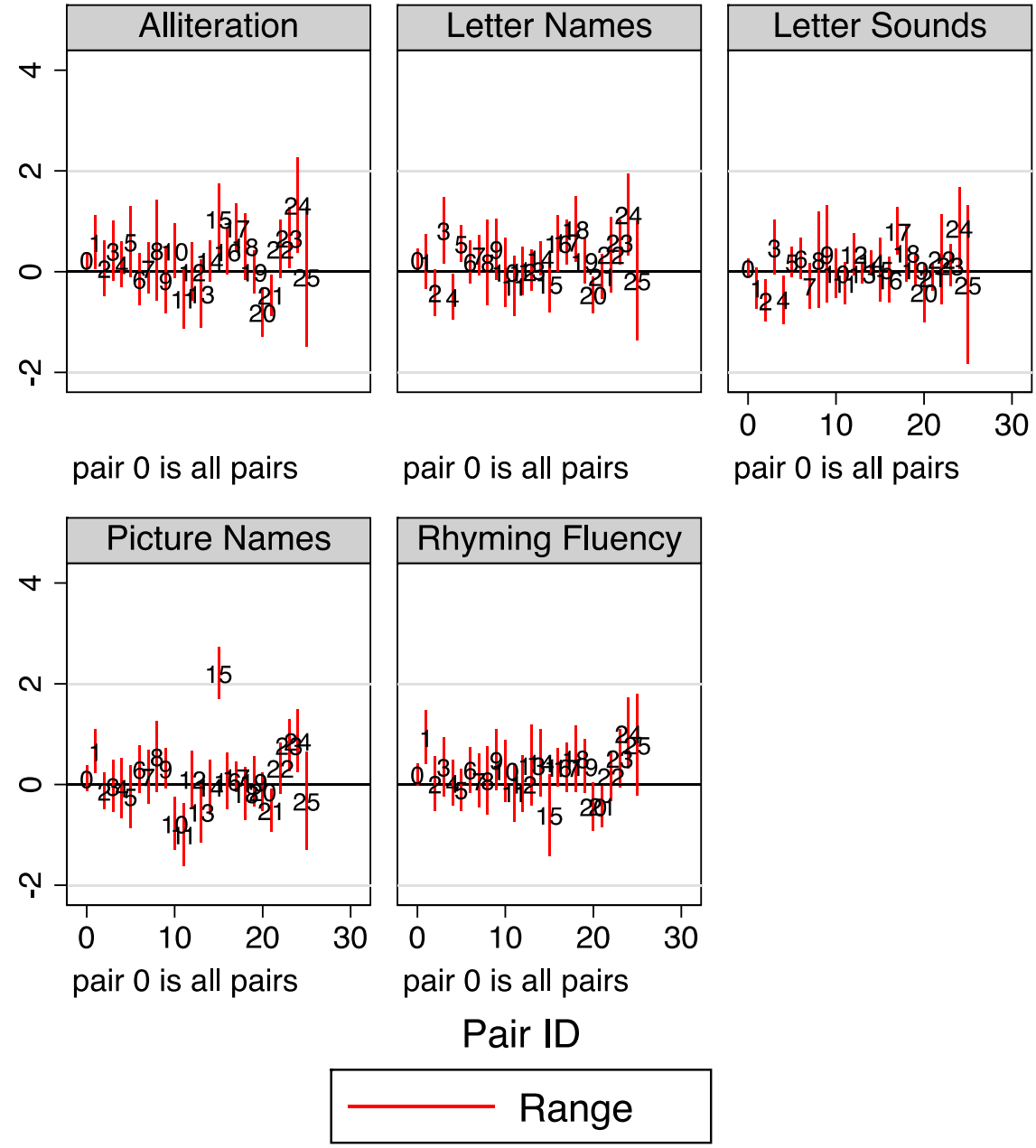


Figure 2: Balance Tests for 4- and 5-year-olds



Summary of Findings. Exhibit A.2.1 below lists the number of site pairs where we found statistically significant imbalance by outcome and age group. The IES What Works Clearinghouse (WWC) considers a standardized difference of 0.25 to be a high level of imbalance, so we also separated our findings by above or below this 0.25 threshold. For example, for alliteration, two sites were imbalanced for 3-year old students, but the difference was less than 0.25 in one of those sites. Alliteration for 4- and 5-year old students was imbalanced for 8 sites; however, the difference was larger than 0.25 for only 5 of those sites.

Exhibit A.2.1: Number of Imbalanced Site Pairs by Outcome and Age Group

Outcome	Number of Imbalanced Pairs					
	Difference < 0.25 Standard Deviations		Difference > 0.25 Standard Deviations		Total	
	3-year old students	4 - and 5- year old students	3- year old students	4 - and 5- year old students	3- year old students	4 - and 5- year old students
Alliteration	1	3	1	5	2	8
Letter Names		3		7		10
Letter Sounds		2		2		4
Picture Names	1	3	2	4	3	7
Rhyming Fluency	1		1	2	2	2

Given the overall number of pairs in the sample (25), the number of pairs where differences greater than 0.25 standard deviations were found is not severe. To account for this moderate imbalance, however, we incorporated the student-level Fall pre-test scores as a covariate in all analyses.

Exhibit A.2.2: Balance Test Results by Site Pair

Pair ID	Age Group	Outcome	Difference	SE(Difference)	DF	Sig.
0	3	Alliteration	0.063	0.104	21	0.548
0	3	Letter Names	0.271	0.134	21	0.056
0	3	Picture Names	0.053	0.178	21	0.768
0	3	Rhyming Fluency	0.225	0.073	21	0.006
0	4	Alliteration	0.237	0.077	50	0.004
0	4	Letter Names	0.239	0.106	50	0.028
0	4	Letter Sounds	0.075	0.089	50	0.407
0	4	Picture Names	0.120	0.130	50	0.361
0	4	Rhyming Fluency	0.206	0.108	50	0.063
1	4	Alliteration	0.583	0.266	73	0.032
1	4	Letter Names	0.193	0.271	72	0.479
1	4	Letter Sounds	-0.320	0.203	73	0.118
1	4	Picture Names	0.660	0.217	72	0.003
1	4	Rhyming Fluency	0.936	0.266	72	0.001
2	4	Alliteration	0.069	0.279	95	0.806
2	4	Letter Names	-0.415	0.230	95	0.074
2	4	Letter Sounds	-0.574	0.210	94	0.008
2	4	Picture Names	-0.129	0.184	95	0.484
2	4	Rhyming Fluency	0.016	0.272	95	0.952
3	3	Alliteration	0.170	0.192	55	0.381
3	3	Letter Names	0.346	0.281	56	0.224
3	3	Picture Names	-0.222	0.271	56	0.415
3	3	Rhyming Fluency	0.143	0.195	56	0.469
3	4	Alliteration	0.415	0.302	135	0.171
3	4	Letter Names	0.821	0.332	135	0.014
3	4	Letter Sounds	0.481	0.274	134	0.081
3	4	Picture Names	-0.030	0.261	135	0.908
3	4	Rhyming Fluency	0.355	0.299	132	0.237
4	3	Alliteration	0.060	0.171	34	0.728
4	3	Letter Names	-0.329	0.236	34	0.173
4	3	Picture Names	0.218	0.355	34	0.542
4	3	Rhyming Fluency	0.237	0.153	34	0.129
4	4	Alliteration	0.145	0.225	50	0.523
4	4	Letter Names	-0.504	0.227	51	0.031
4	4	Letter Sounds	-0.552	0.240	48	0.026
4	4	Picture Names	-0.065	0.297	50	0.827
4	4	Rhyming Fluency	0.029	0.227	51	0.899
5	3	Alliteration	0.356	0.308	39	0.254
5	3	Letter Names	0.199	0.192	39	0.305
5	3	Picture Names	0.047	0.488	41	0.924
5	3	Rhyming Fluency	0.397	0.206	44	0.060
5	4	Alliteration	0.595	0.351	48	0.097
5	4	Letter Names	0.560	0.182	51	0.003
5	4	Letter Sounds	0.192	0.149	48	0.206
5	4	Picture Names	-0.233	0.309	52	0.454
5	4	Rhyming Fluency	-0.099	0.207	51	0.634
6	3	Alliteration	-0.603	0.184	27	0.003

Pair ID	Age Group	Outcome	Difference	SE(Difference)	DF	Sig.
6	3	Letter Names	0.178	0.142	30	0.220
6	3	Picture Names	0.011	0.328	30	0.973
6	3	Rhyming Fluency	-0.215	0.158	30	0.182
6	4	Alliteration	-0.147	0.259	65	0.573
6	4	Letter Names	0.199	0.214	65	0.356
6	4	Letter Sounds	0.266	0.203	64	0.195
6	4	Picture Names	0.306	0.237	65	0.201
6	4	Rhyming Fluency	0.297	0.224	65	0.191
7	3	Alliteration	-0.750	0.546	18	0.186
7	3	Letter Names	0.291	0.198	19	0.157
7	3	Picture Names	0.355	0.388	19	0.372
7	3	Rhyming Fluency	0.411	0.125	19	0.004
7	4	Alliteration	0.069	0.253	63	0.786
7	4	Letter Names	0.329	0.202	65	0.108
7	4	Letter Sounds	-0.281	0.227	63	0.220
7	4	Picture Names	0.162	0.267	65	0.547
7	4	Rhyming Fluency	0.076	0.266	65	0.777
8	3	Alliteration	0.579	0.280	20	0.052
8	3	Letter Names	-0.081	0.155	25	0.605
8	3	Picture Names	0.054	0.365	26	0.883
8	3	Rhyming Fluency	0.231	0.179	26	0.210
8	4	Alliteration	0.426	0.486	25	0.390
8	4	Letter Names	0.190	0.416	28	0.651
8	4	Letter Sounds	0.233	0.465	25	0.621
8	4	Picture Names	0.558	0.346	28	0.118
8	4	Rhyming Fluency	0.086	0.332	28	0.798
9	4	Alliteration	-0.171	0.327	55	0.604
9	4	Letter Names	0.451	0.305	56	0.144
9	4	Letter Sounds	0.348	0.478	54	0.470
9	4	Picture Names	0.320	0.201	56	0.118
9	4	Rhyming Fluency	0.490	0.303	56	0.112
10	4	Alliteration	0.415	0.272	60	0.133
10	4	Letter Names	-0.008	0.341	58	0.982
10	4	Letter Sounds	-0.031	0.244	60	0.898
10	4	Picture Names	-0.775	0.263	58	0.005
10	4	Rhyming Fluency	0.278	0.308	58	0.371
11	4	Alliteration	-0.538	0.289	40	0.070
11	4	Letter Names	-0.284	0.294	43	0.339
11	4	Letter Sounds	-0.230	0.205	40	0.268
11	4	Picture Names	-0.994	0.308	43	0.002
11	4	Rhyming Fluency	-0.158	0.284	43	0.582
12	3	Alliteration	0.431	1.055	4	0.704
12	3	Letter Names	0.309	0.822	4	0.726
12	3	Picture Names	0.072	1.476	4	0.963
12	3	Rhyming Fluency	0.302	1.079	4	0.794
12	4	Alliteration	0.000	0.286	56	0.999
12	4	Letter Names	0.004	0.240	59	0.986
12	4	Letter Sounds	0.349	0.208	56	0.100
12	4	Picture Names	0.107	0.282	59	0.706

Pair ID	Age Group	Outcome	Difference	SE(Difference)	DF	Sig.
12	4	Rhyming Fluency	0.011	0.279	59	0.969
13	4	Alliteration	-0.436	0.328	26	0.195
13	4	Letter Names	0.033	0.198	26	0.867
13	4	Letter Sounds	-0.036	0.097	26	0.717
13	4	Picture Names	-0.543	0.299	26	0.081
13	4	Rhyming Fluency	0.383	0.393	26	0.339
14	3	Alliteration	0.042	0.078	22	0.594
14	3	Letter Names	0.242	0.223	21	0.290
14	3	Picture Names	0.916	0.362	21	0.019
14	3	Rhyming Fluency	-0.012	0.191	21	0.951
14	4	Alliteration	0.219	0.199	29	0.280
14	4	Letter Names	0.265	0.165	31	0.117
14	4	Letter Sounds	0.186	0.111	29	0.104
14	4	Picture Names	-0.049	0.265	31	0.854
14	4	Rhyming Fluency	0.435	0.328	31	0.195
15	4	Alliteration	1.048	0.351	58	0.004
15	4	Letter Names	-0.245	0.277	60	0.379
15	4	Letter Sounds	0.046	0.314	57	0.883
15	4	Picture Names	2.209	0.257	60	0.000
15	4	Rhyming Fluency	-0.604	0.406	60	0.142
16	4	Alliteration	0.400	0.225	54	0.081
16	4	Letter Names	0.565	0.279	56	0.048
16	4	Letter Sounds	-0.158	0.228	53	0.490
16	4	Picture Names	0.071	0.280	56	0.801
16	4	Rhyming Fluency	0.345	0.190	56	0.075
17	4	Alliteration	0.869	0.239	76	0.001
17	4	Letter Names	0.591	0.224	78	0.010
17	4	Letter Sounds	0.804	0.234	76	0.001
17	4	Picture Names	0.150	0.149	78	0.316
17	4	Rhyming Fluency	0.341	0.248	78	0.173
18	3	Alliteration	-0.009	0.137	39	0.947
18	3	Letter Names	0.179	0.197	40	0.367
18	3	Picture Names	-0.851	0.245	41	0.001
18	3	Rhyming Fluency	-0.047	0.144	40	0.749
18	4	Alliteration	0.507	0.332	126	0.130
18	4	Letter Names	0.848	0.326	126	0.010
18	4	Letter Sounds	0.375	0.287	126	0.194
18	4	Picture Names	-0.176	0.265	127	0.508
18	4	Rhyming Fluency	0.519	0.331	126	0.119
19	4	Alliteration	-0.002	0.211	63	0.994
19	4	Letter Names	0.215	0.225	65	0.342
19	4	Letter Sounds	0.040	0.145	63	0.783
19	4	Picture Names	0.057	0.251	65	0.819
19	4	Rhyming Fluency	0.363	0.266	65	0.177
20	4	Alliteration	-0.806	0.246	55	0.002
20	4	Letter Names	-0.453	0.179	54	0.014
20	4	Letter Sounds	-0.407	0.292	55	0.168
20	4	Picture Names	-0.139	0.191	54	0.472
20	4	Rhyming Fluency	-0.434	0.240	54	0.076

Pair ID	Age Group	Outcome	Difference	SE(Difference)	DF	Sig.
21	4	Alliteration	-0.462	0.206	64	0.028
21	4	Letter Names	-0.093	0.222	66	0.676
21	4	Letter Sounds	-0.114	0.133	64	0.394
21	4	Picture Names	-0.513	0.215	66	0.020
21	4	Rhyming Fluency	-0.425	0.215	66	0.052
22	4	Alliteration	0.451	0.291	62	0.126
22	4	Letter Names	0.343	0.374	63	0.362
22	4	Letter Sounds	0.249	0.447	62	0.580
22	4	Picture Names	0.329	0.255	63	0.202
22	4	Rhyming Fluency	0.167	0.244	63	0.497
23	3	Alliteration	0.761	0.340	20	0.036
23	3	Letter Names	0.365	0.399	20	0.371
23	3	Picture Names	1.034	0.381	20	0.013
23	3	Rhyming Fluency	0.565	0.373	20	0.145
23	4	Alliteration	0.669	0.300	54	0.030
23	4	Letter Names	0.583	0.220	56	0.011
23	4	Letter Sounds	0.123	0.211	54	0.561
23	4	Picture Names	0.781	0.255	56	0.003
23	4	Rhyming Fluency	0.524	0.286	56	0.073
24	4	Alliteration	1.323	0.463	29	0.008
24	4	Letter Names	1.134	0.400	35	0.008
24	4	Letter Sounds	0.865	0.396	30	0.037
24	4	Picture Names	0.863	0.308	35	0.008
24	4	Rhyming Fluency	1.015	0.346	34	0.006
25	3	Alliteration	0.652	0.913	18	0.484
25	3	Letter Names	-0.682	1.036	18	0.519
25	3	Picture Names	1.210	1.030	18	0.255
25	3	Rhyming Fluency	0.893	1.002	18	0.385
25	4	Alliteration	-0.103	0.673	28	0.880
25	4	Letter Names	-0.175	0.580	29	0.765
25	4	Letter Sounds	-0.258	0.766	27	0.739
25	4	Picture Names	-0.325	0.474	29	0.498
25	4	Rhyming Fluency	0.790	0.493	29	0.120

Appendix A.3: Detailed Methodology

The goal of this analysis is to estimate a plausible effect of the Minnesota Reading Corps program, which is the average treatment effect within each site pair. To accomplish this goal we employ an econometric fixed effects model. Also known as a within-estimator, this model is ideal for two reasons. First, it removes all effects on the outcomes associated with the site pair (namely, all the characteristics used to match the sites). Second, it also allows for a correlation between the predictors and the fixed site pair effects. That is, it allows for the site pair average Fall score to be correlated with the average Spring score.
The fixed effects model achieves this by a process of subtracting the group means for each variable. Each variable (dependent or independent) is transformed by subtracting the site pair mean and adding the overall mean. Thus, for student *i* in pair *j*

$$x_{ij}^{*} = x_{ij} - \bar{x}_j + \bar{x}.$$

The drawback of the fixed effects approach is that it does not allow for pair level covariates. That is, variables used to match the pairs that are common between the sites cannot be entered as covariates because $x_{ij}^{*} = x_{ij} - \bar{x}_j + \bar{x}$ evaluates to the constant overall mean and cannot be used as a predictor. In order to use pair-level covariates, we would need to use a random effects model. However, a random effects model depends on the assumption that the correlation between the predictors and the fixed site pair effects is 0. In order to validate this approach, a Hausman test must confirm that the fixed effects and random effects models produce similar within-pair estimates.

We performed tests to compare the findings from the fixed effects model to those of a random effects model. For most outcome measures, the effect of age was sufficiently different between the two estimators that a traditional Hausman test could not be conducted. However in the case of letter names, the random effects model's results were inconsistent from the fixed effects model to raise questions of the validity of the random effects model. As a result, we chose to employ the fixed effects approach to estimate the average within-pair treatment effect.

Coding Procedures

The general modeling is an ANCOVA approach, where the dependent variable (Winter or Spring score) is regressed onto the Fall score and treatment (Minnesota Reading Corps) indicator. The treatment indicator coefficient is the average difference between the dependent score (Winter or Spring) and the Fall score, controlling for the Fall score (baseline). For the 4- and 5-year old student analyses, an additional indicator for 5-year old students (vs. 4-year-old students) is employed. All scores are centered on the Fall mean scores to allow the intercept to be interpreted as the comparison mean difference. The regression model is in the following form, for the *ith* student in school *j* in pair *k*

$$y_{ijk} - \bar{x} = b_0 + b_1 MRC_{jk} + b_3\left(x_{ijk} - \bar{x}\right) + b_4 Age_{ijk} + e_{ijk}.$$

Note that the dependent variable, *y*, which is the Spring or Winter score, is coded as a deviation from the mean of the covariate *x*, which is the Fall score. The pretest covariate, *x*, is also the deviation from the mean of *x*. This coding allows for the intercept, b_0, to equal the *average* deviation of *y* from the mean of *x* for comparison students with an *average* value of *x*. In other words, the intercept is the average growth for a student with an average Fall score. The effect, b_1, is the *difference* in the *growth* for a student with an average Fall score associated with the Minnesota Reading Corps program. The effect, b_4, is simply an age adjustment for the models that include both 4- and 5-year-old students.

Effect size calculations

A typical estimation of a mean difference effect size is Cohen's *d*

$$d = \frac{m_A - m_B}{s}$$

which is the difference in the mean outcome divided by a standard deviation. In this case, the difference between a treatment and comparison student is simply the Minnesota Reading Corps coefficient (holding the Fall score and mean and age constant).

$$\begin{aligned} MRC\_Effect &= \left(b_0 + b_1 1 + b_3\left(x_{ij} - \bar{x}\right)\right) - \left(b_0 + b_1 0 + b_3\left(x_{ij} - \bar{x}\right)\right) \\ &= \left(b_0 + b_1 1 + b_3 0\right) - \left(b_0 + b_1 0 + b_3 0\right) \\ &= \left(b_0 + b_1\right) - \left(b_0\right) \\ &= b_1 \end{aligned}$$

However, the choice of standard deviation is more challenging. The procedure we employed to calculate the standard deviation was to estimate the within-age group variance of each subject's Fall score and take the square root of this variance as the standard deviation. That is, given a set of q = 1, 2, … m ages, the standard deviation is the standard deviation of the residual differences of the outcome from the age-specific means

$$s_y = \sqrt{\frac{\sum_{m}\sum_{n_q}\left(y_{iq} - \bar{y}_q\right)^2}{N - 1}}$$

Using this standard deviation, our effect sizes are equal to

$$d_y = \frac{b_{1y}}{s_y}$$

Using a common standard deviation for each outcome allows the effect sizes to be comparable across ages and subgroups.

Power analysis

Our power analysis estimated the minimal detectable effect[3] from a two-level cluster randomized trial[4] using design parameters derived from previous benchmark outcome data for the 2012-13 school year recorded by the Minnesota Reading Corps program. This formula is[5]:

$$MDES \approx M_{J-K}\sqrt{\frac{r\left(1 - R_2^2\right)}{J/4} + \frac{\left(1 - r\right)\left(1 - R_1^2\right)}{\left(nJ\right)/4}}$$

where J is the total number of schools, K is the number of predictors (1), r is the intraclass correlation (ICC), R_2^2 is the effect of a pretest at the school level, R_1^2 is the effect of a pretest at the student level, n is the number of students per school, and M is a factor relating to the t-distributions values for alpha (0.05) and power (0.80), which in this case is roughly equal to 2.8.

[3] Bloom, H. S. (1995). Minimum detectable effects a simple way to report the statistical power of experimental designs. *Evaluation review*, *19*(5), 547-556.

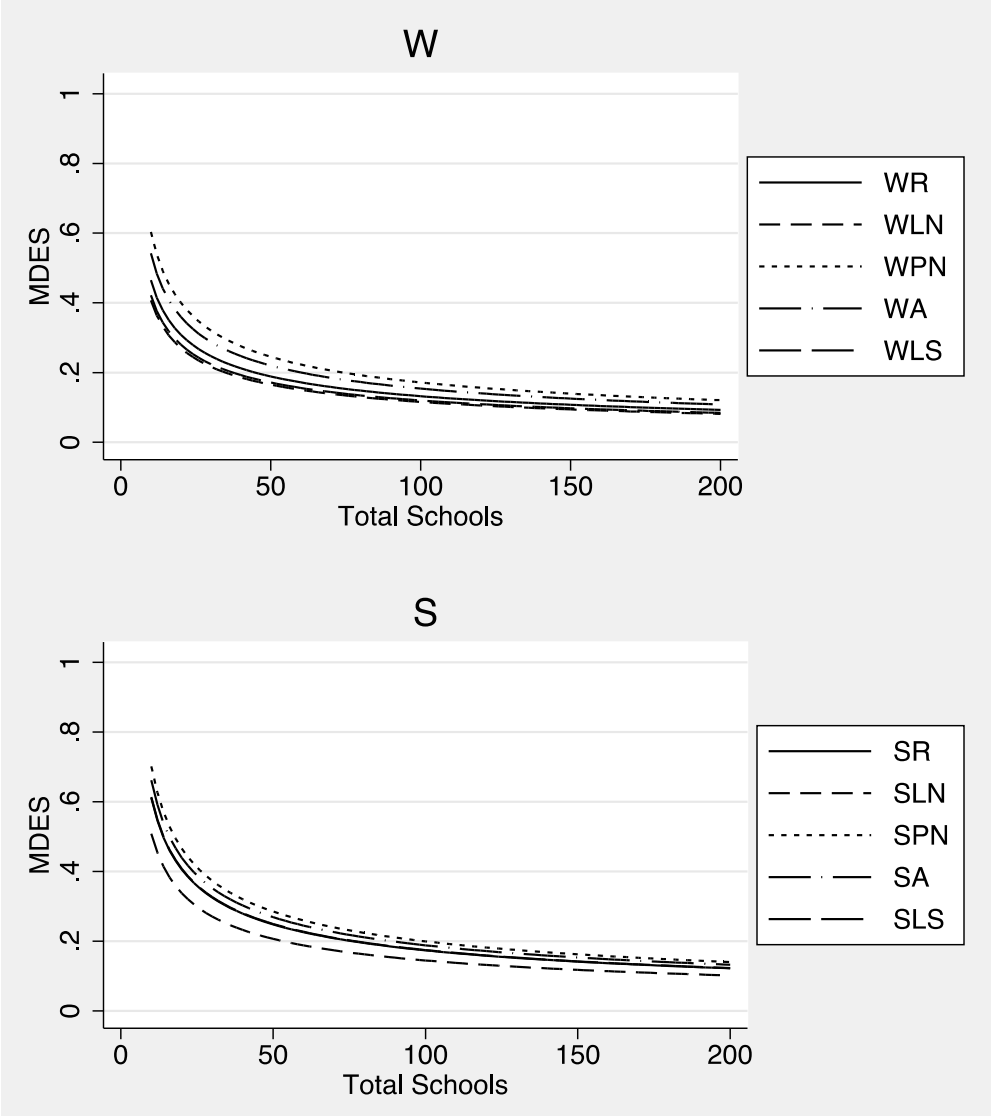
[4] Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, *29*(1), 30-59.

[5] Assumes equal allocation to treatment and comparison in expression 4 in Bloom, et al. 2007

Our power analysis employed the following parameters:

		ICC	R-square 2	R-square 1
Winter Outcome	Abbreviation			
Rhyming	W R	0.10	0.66	0.37
Letter names	W LN	0.12	0.76	0.58
Picture names	W PN	0.19	0.61	0.29
Alliteration	W A	0.11	0.51	0.25
Letter sounds	W LS	0.13	0.76	0.47
Spring Outcome				
Rhyming	S R	0.12	0.39	0.28
Letter names	S LN	0.09	0.43	0.46
Picture names	S PN	0.21	0.49	0.31
Alliteration	S A	0.16	0.45	0.20
Letter sounds	S LS	0.13	0.43	0.33

With picture names having the largest intraclass correlation. In the figure below, we use the 2012-13 Spring scores to estimate that 25 schools per condition (50 total) is sufficient to detect an effect size of about 0.2 for the literacy outcomes, assuming 30 students per school.


W
MDES
0
.2
.4
.6
.8
1
0
50
100
150
200
Total Schools
WR
WLN
WPN
WA
WLS
S
MDES
0
.2
.4
.6
.8
1
0
50
100
150
200
Total Schools
SR
SLN
SPN
SA
SLS

Appendix B.1: Fall-Winter Models and Effects Tables

Table B.1.1: ANCOVA Regression Models Predicting Winter Letter Sound Score of 4-5-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	0.80	(0.56)
Fall Score	1.07 ***	(0.06)
Age 5	0.79 *	(0.36)
Intercept	2.74 ***	(0.41)
Model with gender moderation		
Treatment vs. Comparison	0.19	(0.54)
Female	-0.44	(0.66)
Treatment vs. Comparison x Female	1.29	(0.69)
Fall Score	1.07 ***	(0.06)
Age 5	0.79 *	(0.34)
Intercept	2.95 ***	(0.45)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	0.97	(0.58)
Dual language vs. single language	1.46	(1.18)
Treatment vs. Comparison x dual language	-1.15	(1.86)
Fall Score	1.07 ***	(0.07)
Age 5	0.80 *	(0.38)
Intercept	2.53 ***	(0.44)
Model with White/non-White moderation		
Treatment vs. Comparison	1.02	(0.68)
non-White vs. White	0.55	(0.94)
Treatment vs. Comparison x non-White	-0.96	(1.35)
Fall Score	1.06 ***	(0.07)
Age 5	0.79 *	(0.38)
Intercept	2.62 ***	(0.50)
Model with Public School students		
Treatment vs. Comparison	0.75	(0.68)
Fall Score	1.08 ***	(0.08)
Age 5	0.90 *	(0.40)
Intercept	3.08 ***	(0.49)
Model with Head Start students		
Treatment vs. Comparison	1.06	(0.95)
Fall Score	0.97 ***	(0.14)
Age 5	-0.24	(0.73)
Intercept	0.94	(0.90)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.1.2: ANCOVA Regression Models Predicting Winter Rhyming Score of 4-5-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	1.64 *	(0.59)
Fall Score	0.72 ***	(0.06)
Age 5	1.08 *	(0.46)
Intercept	2.31 ***	(0.38)
Model with gender moderation		
Treatment vs. Comparison	1.70 *	(0.70)
Female	0.33	(0.53)
Treatment vs. Comparison x Female	-0.11	(0.53)
Fall Score	0.72 ***	(0.06)
Age 5	1.12 *	(0.48)
Intercept	2.13 ***	(0.52)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	1.65 *	(0.65)
Dual language vs. single language	-0.86	(1.19)
Treatment vs. Comparison x dual language	-0.08	(1.19)
Fall Score	0.72 ***	(0.06)
Age 5	1.08 *	(0.47)
Intercept	2.42 ***	(0.43)
Model with White/non-White moderation		
Treatment vs. Comparison	1.47	(0.72)
non-White vs. White	-1.88 **	(0.61)
Treatment vs. Comparison x non-White	0.42	(0.78)
Fall Score	0.71 ***	(0.06)
Age 5	1.08 *	(0.47)
Intercept	2.87 ***	(0.46)
Model with Public School students		
Treatment vs. Comparison	1.77 *	(0.76)
Fall Score	0.73 ***	(0.07)
Age 5	1.21 *	(0.49)
Intercept	2.36 ***	(0.47)
Model with Head Start students		
Treatment vs. Comparison	1.79 *	(0.50)
Fall Score	0.73 ***	(0.05)
Age 5	-0.85	(2.36)
Intercept	1.51 **	(0.37)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.1.3: ANCOVA Regression Models Predicting Winter Letter Name Score of 4-5-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	1.83 *	(0.87)
Fall Score	0.98 ***	(0.04)
Age 5	1.62 **	(0.55)
Intercept	6.08 ***	(0.56)
Model with gender moderation		
Treatment vs. Comparison	2.19 *	(0.89)
Female	1.70 *	(0.80)
Treatment vs. Comparison x Female	-0.63	(0.95)
Fall Score	0.98 ***	(0.04)
Age 5	1.77 **	(0.59)
Intercept	5.21 ***	(0.67)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	2.04 *	(0.89)
Dual language vs. single language	2.06	(1.42)
Treatment vs. Comparison x dual language	-1.35	(1.55)
Fall Score	0.98 ***	(0.04)
Age 5	1.65 **	(0.55)
Intercept	5.77 ***	(0.60)
Model with White/non-White moderation		
Treatment vs. Comparison	2.06	(1.04)
non-White vs. White	0.77	(1.23)
Treatment vs. Comparison x non-White	-1.18	(1.55)
Fall Score	0.98 ***	(0.05)
Age 5	1.37 *	(0.54)
Intercept	5.97 ***	(0.74)
Model with Public School students		
Treatment vs. Comparison	1.54	(1.06)
Fall Score	0.96 ***	(0.05)
Age 5	1.60 *	(0.60)
Intercept	6.55 ***	(0.70)
Model with Head Start students		
Treatment vs. Comparison	3.14	(1.65)
Fall Score	1.13 ***	(0.05)
Age 5	2.99	(2.67)
Intercept	4.72 **	(1.16)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.1.4: ANCOVA Regression Models Predicting Winter Picture Name Score of 4-5-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	3.17 ***	(0.65)
Fall Score	0.50 ***	(0.03)
Age 5	0.19	(0.39)
Intercept	2.67 ***	(0.42)
Model with gender moderation		
Treatment vs. Comparison	2.94 **	(0.80)
Female	-0.29	(0.51)
Treatment vs. Comparison x Female	0.48	(0.60)
Fall Score	0.50 ***	(0.03)
Age 5	0.18	(0.38)
Intercept	2.81 ***	(0.50)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	2.85 ***	(0.70)
Dual language vs. single language	-3.04 *	(1.22)
Treatment vs. Comparison x dual language	2.14	(1.82)
Fall Score	0.49 ***	(0.03)
Age 5	0.19	(0.38)
Intercept	3.13 ***	(0.48)
Model with White/non-White moderation		
Treatment vs. Comparison	2.79 ***	(0.74)
non-White vs. White	-3.46 **	(1.13)
Treatment vs. Comparison x non-White	2.10	(1.24)
Fall Score	0.48 ***	(0.04)
Age 5	0.27	(0.38)
Intercept	3.44 ***	(0.56)
Model with Public School students		
Treatment vs. Comparison	3.31 **	(0.83)
Fall Score	0.50 ***	(0.04)
Age 5	0.05	(0.42)
Intercept	2.79 ***	(0.53)
Model with Head Start students		
Treatment vs. Comparison	2.41 *	(0.89)
Fall Score	0.50 ***	(0.08)
Age 5	1.80	(0.74)
Intercept	2.42 **	(0.53)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.1.5: ANCOVA Regression Models Predicting Winter Alliteration Score of 4-5-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	1.54 ***	(0.38)
Fall Score	0.69 ***	(0.05)
Age 5	0.77 *	(0.35)
Intercept	1.05 **	(0.29)
Model with gender moderation		
Treatment vs. Comparison	1.41 *	(0.55)
Female	0.24	(0.44)
Treatment vs. Comparison x Female	0.29	(0.58)
Fall Score	0.68 ***	(0.05)
Age 5	0.82 *	(0.34)
Intercept	0.93 *	(0.41)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	1.61 ***	(0.41)
Dual language vs. single language	-0.27	(0.71)
Treatment vs. Comparison x dual language	-0.48	(0.93)
Fall Score	0.68 ***	(0.05)
Age 5	0.77 *	(0.35)
Intercept	1.09 **	(0.32)
Model with White/non-White moderation		
Treatment vs. Comparison	1.45 **	(0.44)
non-White vs. White	-0.23	(0.73)
Treatment vs. Comparison x non-White	-0.15	(0.72)
Fall Score	0.69 ***	(0.05)
Age 5	0.72	(0.36)
Intercept	1.22 **	(0.36)
Model with Public School students		
Treatment vs. Comparison	1.63 **	(0.45)
Fall Score	0.72 ***	(0.06)
Age 5	0.64	(0.37)
Intercept	1.25 **	(0.34)
Model with Head Start students		
Treatment vs. Comparison	1.21	(0.77)
Fall Score	0.51 **	(0.13)
Age 5	0.76	(1.15)
Intercept	0.10	(0.57)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.1.6: ANCOVA Regression Models Predicting Winter Rhyming Score of 3-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	0.64	(0.49)
Fall Score	0.80***	(0.08)
Intercept	0.40	(0.47)
Model with gender moderation		
Treatment vs. Comparison	1.25*	(0.47)
Female	0.92	(0.65)
Treatment vs. Comparison x Female	-1.23	(0.74)
Fall Score	0.80***	(0.08)
Intercept	-0.07	(0.46)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	0.82	(0.41)
Dual language vs. single language	-1.28**	(0.39)
Treatment vs. Comparison x dual language	-0.60	(0.59)
Fall Score	0.78***	(0.08)
Intercept	0.54	(0.41)
Model with White/non-White moderation		
Treatment vs. Comparison	1.00	(0.59)
non-White vs. White	0.40	(0.91)
Treatment vs. Comparison x non-White	-0.88	(0.96)
Fall Score	0.80***	(0.08)
Intercept	0.25	(0.57)
Model with Public School students		
Treatment vs. Comparison	1.10	(0.55)
Fall Score	0.89**	(0.11)
Intercept	0.36	(0.48)
Model with Head Start students		
Treatment vs. Comparison	0.25	(0.78)
Fall Score	0.67*	(0.20)
Intercept	-0.15	(1.15)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.1.7: ANCOVA Regression Models Predicting Winter Letter Name Score of 3-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	0.41	(0.77)
Fall Score	0.93 ***	(0.06)
Intercept	3.17 ***	(0.54)
Model with gender moderation		
Treatment vs. Comparison	0.50	(1.15)
Female	0.53	(1.16)
Treatment vs. Comparison x Female	-0.14	(1.59)
Fall Score	0.93 ***	(0.05)
Intercept	2.90 *	(0.92)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	0.78	(0.79)
Dual language vs. single language	3.44	(2.09)
Treatment vs. Comparison x dual language	-2.80	(2.91)
Fall Score	0.93 ***	(0.06)
Intercept	2.70 ***	(0.58)
Model with White/non-White moderation		
Treatment vs. Comparison	0.09	(0.68)
non-White vs. White	0.29	(1.97)
Treatment vs. Comparison x non-White	1.14	(2.18)
Fall Score	0.94 ***	(0.06)
Intercept	3.01 **	(0.68)
Model with Public School students		
Treatment vs. Comparison	-0.96	(0.33)
Fall Score	0.94 ***	(0.04)
Intercept	3.68 ***	(0.23)
Model with Head Start students		
Treatment vs. Comparison	1.48	(1.04)
Fall Score	1.08 ***	(0.09)
Intercept	3.42 ***	(0.44)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.1.8: ANCOVA Regression Models Predicting Winter Picture Name Score of 3-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	0.41	(0.77)
Fall Score	0.93***	(0.06)
Intercept	3.17***	(0.54)
Model with gender moderation		
Treatment vs. Comparison	0.50	(1.15)
Female	0.53	(1.16)
Treatment vs. Comparison x Female	-0.14	(1.59)
Fall Score	0.93***	(0.05)
Intercept	2.90*	(0.92)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	0.78	(0.79)
Dual language vs. single language	3.44	(2.09)
Treatment vs. Comparison x dual language	-2.80	(2.91)
Fall Score	0.93***	(0.06)
Intercept	2.70***	(0.58)
Model with White/non-White moderation		
Treatment vs. Comparison	0.09	(0.68)
non-White vs. White	0.29	(1.97)
Treatment vs. Comparison x non-White	1.14	(2.18)
Fall Score	0.94***	(0.06)
Intercept	3.01**	(0.68)
Model with Public School students		
Treatment vs. Comparison	-0.96	(0.33)
Fall Score	0.94***	(0.04)
Intercept	3.68***	(0.23)
Model with Head Start students		
Treatment vs. Comparison	1.48	(1.04)
Fall Score	1.08***	(0.09)
Intercept	3.42***	(0.44)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.1.9: ANCOVA Regression Models Predicting Winter Alliteration Score of 3-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	1.43	(0.68)
Fall Score	0.38	(0.19)
Intercept	0.54	(0.66)
Model with gender moderation		
Treatment vs. Comparison	1.67	(0.86)
Female	0.81	(0.63)
Treatment vs. Comparison x Female	-0.30	(0.91)
Fall Score	0.39	(0.19)
Intercept	0.09	(0.76)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	1.30	(0.74)
Dual language vs. single language	-1.65 *	(0.70)
Treatment vs. Comparison x dual language	1.26	(0.80)
Fall Score	0.38	(0.19)
Intercept	0.72	(0.70)
Model with White/non-White moderation		
Treatment vs. Comparison	0.92	(0.64)
non-White vs. White	-0.79	(0.81)
Treatment vs. Comparison x non-White	1.41	(0.74)
Fall Score	0.39	(0.20)
Intercept	0.79	(0.65)
Model with Public School students		
Treatment vs. Comparison	-0.18	(0.67)
Fall Score	0.49	(0.25)
Intercept	1.03	(0.72)
Model with Head Start students		
Treatment vs. Comparison	2.62 *	(0.70)
Fall Score	0.14	(0.16)
Intercept	-0.34	(0.63)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Appendix B.2: Fall-Spring Models and Effects Tables

Table B.2.1: ANCOVA Regression Models Predicting Spring Letter Sound Score of 4-5-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	3.80 ***	(0.90)
Fall Score	1.16 ***	(0.08)
Age 5	0.19	(0.58)
Intercept	4.42 ***	(0.58)
Model with gender moderation		
Treatment vs. Comparison	3.31 **	(0.89)
Female	0.37	(0.76)
Treatment vs. Comparison x Female	1.04	(0.91)
Fall Score	1.16 ***	(0.08)
Age 5	0.35	(0.54)
Intercept	4.23 ***	(0.63)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	3.73 ***	(0.83)
Dual language vs. single language	-1.24	(1.67)
Treatment vs. Comparison x dual language	1.14	(2.31)
Fall Score	1.15 ***	(0.08)
Age 5	0.17	(0.58)
Intercept	4.54 ***	(0.56)
Model with White/non-White moderation		
Treatment vs. Comparison	3.51 ***	(0.87)
non-White vs. White	-0.10	(1.48)
Treatment vs. Comparison x non-White	0.01	(2.17)
Fall Score	1.15 ***	(0.08)
Age 5	0.19	(0.58)
Intercept	4.65 ***	(0.59)
Model with Public School students		
Treatment vs. Comparison	3.58 **	(1.13)
Fall Score	1.17 ***	(0.10)
Age 5	0.29	(0.61)
Intercept	4.86 ***	(0.68)
Model with Head Start students		
Treatment vs. Comparison	4.93 **	(1.28)
Fall Score	1.09 *	(0.31)
Age 5	-1.44 *	(0.53)
Intercept	2.14	(1.32)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.2.2: ANCOVA Regression Models Predicting Spring Rhyming Score of 4-5-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	3.51 ***	(0.52)
Fall Score	0.68 ***	(0.05)
Age 5	1.19	(0.60)
Intercept	3.77 ***	(0.31)
Model with gender moderation		
Treatment vs. Comparison	3.62 ***	(0.71)
Female	1.10	(0.59)
Treatment vs. Comparison x Female	-0.16	(0.78)
Fall Score	0.68 ***	(0.05)
Age 5	1.36 *	(0.61)
Intercept	3.22 ***	(0.50)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	3.52 ***	(0.57)
Dual language vs. single language	-2.66 **	(0.82)
Treatment vs. Comparison x dual language	1.16	(1.08)
Fall Score	0.67 ***	(0.05)
Age 5	1.16	(0.61)
Intercept	4.03 ***	(0.36)
Model with White/non-White moderation		
Treatment vs. Comparison	3.32 ***	(0.70)
non-White vs. White	-2.45 **	(0.85)
Treatment vs. Comparison x non-White	1.07	(0.93)
Fall Score	0.66 ***	(0.05)
Age 5	1.30 *	(0.62)
Intercept	4.34 ***	(0.44)
Model with Public School students		
Treatment vs. Comparison	3.74 ***	(0.67)
Fall Score	0.68 ***	(0.06)
Age 5	1.38 *	(0.64)
Intercept	3.66 ***	(0.39)
Model with Head Start students		
Treatment vs. Comparison	3.35 **	(0.83)
Fall Score	0.73 ***	(0.05)
Age 5	-0.71	(1.05)
Intercept	3.79 ***	(0.56)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.2.3: ANCOVA Regression Models Predicting Spring Letter Name Score of 4-5-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	4.15 **	(1.17)
Fall Score	1.00 ***	(0.04)
Age 5	0.29	(0.50)
Intercept	9.87 ***	(0.72)
Model with gender moderation		
Treatment vs. Comparison	4.36 **	(1.20)
Female	2.19 *	(0.92)
Treatment vs. Comparison x Female	-0.39	(1.04)
Fall Score	1.00 ***	(0.04)
Age 5	0.55	(0.50)
Intercept	8.78 ***	(0.85)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	4.14 **	(1.20)
Dual language vs. single language	-0.71	(2.70)
Treatment vs. Comparison x dual language	0.33	(2.93)
Fall Score	1.00 ***	(0.04)
Age 5	0.28	(0.50)
Intercept	9.94 ***	(0.78)
Model with White/non-White moderation		
Treatment vs. Comparison	3.18 *	(1.40)
non-White vs. White	-2.18	(1.43)
Treatment vs. Comparison x non-White	2.52	(1.86)
Fall Score	0.99 ***	(0.04)
Age 5	0.13	(0.54)
Intercept	10.72 ***	(0.94)
Model with Public School students		
Treatment vs. Comparison	3.47 *	(1.46)
Fall Score	0.99 ***	(0.05)
Age 5	0.32	(0.56)
Intercept	10.61 ***	(0.89)
Model with Head Start students		
Treatment vs. Comparison	6.18 *	(2.11)
Fall Score	1.14 ***	(0.03)
Age 5	-0.07	(1.93)
Intercept	8.32 ***	(1.27)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.2.4: ANCOVA Regression Models Predicting Spring Picture Name Score of 4-5-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	4.03***	(0.66)
Fall Score	0.46***	(0.03)
Age 5	-0.28	(0.44)
Intercept	4.65***	(0.39)
Model with gender moderation		
Treatment vs. Comparison	3.49***	(0.88)
Female	-0.89	(0.76)
Treatment vs. Comparison x Female	1.11	(0.82)
Fall Score	0.46***	(0.03)
Age 5	-0.34	(0.46)
Intercept	5.09***	(0.60)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	3.81***	(0.69)
Dual language vs. single language	-4.37***	(0.98)
Treatment vs. Comparison x dual language	3.65**	(1.28)
Fall Score	0.44***	(0.03)
Age 5	-0.30	(0.45)
Intercept	5.13***	(0.43)
Model with White/non-White moderation		
Treatment vs. Comparison	3.59***	(0.75)
non-White vs. White	-3.40*	(1.22)
Treatment vs. Comparison x non-White	2.26	(1.22)
Fall Score	0.44***	(0.03)
Age 5	-0.03	(0.39)
Intercept	5.40***	(0.50)
Model with Public School students		
Treatment vs. Comparison	4.05***	(0.82)
Fall Score	0.47***	(0.03)
Age 5	-0.44	(0.45)
Intercept	4.78***	(0.45)
Model with Head Start students		
Treatment vs. Comparison	4.91**	(0.93)
Fall Score	0.40***	(0.06)
Age 5	1.56	(1.28)
Intercept	3.52**	(0.61)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.2.5: ANCOVA Regression Models Predicting Spring Alliteration Score of 4-5-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	2.68 ***	(0.44)
Fall Score	0.70 ***	(0.06)
Age 5	0.41	(0.57)
Intercept	2.68 ***	(0.28)
Model with gender moderation		
Treatment vs. Comparison	3.00 ***	(0.42)
Female	1.50 *	(0.59)
Treatment vs. Comparison x Female	-0.69	(0.79)
Fall Score	0.69 ***	(0.06)
Age 5	0.62	(0.57)
Intercept	1.97 ***	(0.27)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	2.86 ***	(0.43)
Dual language vs. single language	-0.60	(1.15)
Treatment vs. Comparison x dual language	-1.18	(1.49)
Fall Score	0.69 ***	(0.06)
Age 5	0.39	(0.58)
Intercept	2.71 ***	(0.29)
Model with White/non-White moderation		
Treatment vs. Comparison	2.80 ***	(0.55)
non-White vs. White	-0.81	(0.73)
Treatment vs. Comparison x non-White	-0.24	(0.80)
Fall Score	0.69 ***	(0.06)
Age 5	0.37	(0.60)
Intercept	2.84 ***	(0.36)
Model with Public School students		
Treatment vs. Comparison	2.43 ***	(0.49)
Fall Score	0.71 ***	(0.07)
Age 5	0.46	(0.63)
Intercept	3.03 ***	(0.29)
Model with Head Start students		
Treatment vs. Comparison	3.65 *	(1.13)
Fall Score	0.57 *	(0.18)
Age 5	-0.59	(0.44)
Intercept	1.13	(0.88)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.2.6: ANCOVA Regression Models Predicting Spring Rhyming Score of 3-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	2.30***	(0.37)
Fall Score	0.65***	(0.14)
Intercept	0.86	(0.48)
Model with gender moderation		
Treatment vs. Comparison	2.98**	(0.68)
Female	0.70	(0.95)
Treatment vs. Comparison x Female	-1.43	(0.97)
Fall Score	0.64***	(0.14)
Intercept	0.49	(0.84)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	2.30***	(0.44)
Dual language vs. single language	-1.83*	(0.68)
Treatment vs. Comparison x dual language	0.49	(1.16)
Fall Score	0.63***	(0.14)
Intercept	1.09	(0.52)
Model with White/non-White moderation		
Treatment vs. Comparison	1.95**	(0.61)
non-White vs. White	-0.78	(0.51)
Treatment vs. Comparison x non-White	0.84	(0.82)
Fall Score	0.65**	(0.14)
Intercept	1.16*	(0.40)
Model with Public School students		
Treatment vs. Comparison	2.00	(0.73)
Fall Score	0.88*	(0.18)
Intercept	2.05	(0.81)
Model with Head Start students		
Treatment vs. Comparison	2.46**	(0.38)
Fall Score	0.49*	(0.18)
Intercept	-0.38	(0.55)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.2.7: ANCOVA Regression Models Predicting Spring Letter Name Score of 3-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	1.41	(1.41)
Fall Score	0.80 ***	(0.07)
Intercept	5.11 ***	(1.02)
Model with gender moderation		
Treatment vs. Comparison	2.48	(1.59)
Female	2.43	(1.42)
Treatment vs. Comparison x Female	-2.00	(1.70)
Fall Score	0.80 ***	(0.07)
Intercept	3.84 *	(1.35)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	1.37	(1.44)
Dual language vs. single language	0.43	(2.18)
Treatment vs. Comparison x dual language	0.22	(2.86)
Fall Score	0.80 ***	(0.07)
Intercept	5.04 **	(1.10)
Model with White/non-White moderation		
Treatment vs. Comparison	0.27	(0.80)
non-White vs. White	-2.79	(1.62)
Treatment vs. Comparison x non-White	3.40	(2.52)
Fall Score	0.80 ***	(0.07)
Intercept	6.11 ***	(0.69)
Model with Public School students		
Treatment vs. Comparison	-0.18	(1.55)
Fall Score	0.87 ***	(0.05)
Intercept	5.38 *	(1.08)
Model with Head Start students		
Treatment vs. Comparison	2.60	(2.35)
Fall Score	0.89 ***	(0.09)
Intercept	5.24 *	(1.49)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.2.8: ANCOVA Regression Models Predicting Spring Picture Name Score of 3-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	3.52 **	(1.06)
Fall Score	0.56 ***	(0.07)
Intercept	3.43 **	(0.88)
Model with gender moderation		
Treatment vs. Comparison	2.18	(1.23)
Female	-2.10	(1.56)
Treatment vs. Comparison x Female	2.88	(2.05)
Fall Score	0.56 ***	(0.07)
Intercept	4.42 **	(0.97)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	3.73 **	(1.16)
Dual language vs. single language	-3.79	(2.88)
Treatment vs. Comparison x dual language	-2.33	(2.57)
Fall Score	0.49 ***	(0.07)
Intercept	3.89 ***	(0.82)
Model with White/non-White moderation		
Treatment vs. Comparison	3.20 *	(1.14)
non-White vs. White	-2.21	(1.59)
Treatment vs. Comparison x non-White	0.75	(1.91)
Fall Score	0.54 ***	(0.07)
Intercept	4.23 ***	(0.82)
Model with Public School students		
Treatment vs. Comparison	2.43	(1.32)
Fall Score	0.69 **	(0.11)
Intercept	3.91 **	(0.63)
Model with Head Start students		
Treatment vs. Comparison	5.40 *	(1.95)
Fall Score	0.45 **	(0.08)
Intercept	1.81	(1.75)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Table B.2.9: ANCOVA Regression Models Predicting Spring Alliteration Score of 3-year-olds

Model with all students	Estimate	Standard Error
Treatment vs. Comparison	3.52 **	(1.06)
Fall Score	0.56 ***	(0.07)
Intercept	3.43 **	(0.88)
Model with gender moderation		
Treatment vs. Comparison	2.18	(1.23)
Female	-2.10	(1.56)
Treatment vs. Comparison x Female	2.88	(2.05)
Fall Score	0.56 ***	(0.07)
Intercept	4.42 **	(0.97)
Model with Dual Language Learner moderation		
Treatment vs. Comparison	3.73 **	(1.16)
Dual language vs. single language	-3.79	(2.88)
Treatment vs. Comparison x dual language	-2.33	(2.57)
Fall Score	0.49 ***	(0.07)
Intercept	3.89 ***	(0.82)
Model with White/non-White moderation		
Treatment vs. Comparison	3.20 *	(1.14)
non-White vs. White	-2.21	(1.59)
Treatment vs. Comparison x non-White	0.75	(1.91)
Fall Score	0.54 ***	(0.07)
Intercept	4.23 ***	(0.82)
Model with Public School students		
Treatment vs. Comparison	2.43	(1.32)
Fall Score	0.69 **	(0.11)
Intercept	3.91 **	(0.63)
Model with Head Start students		
Treatment vs. Comparison	5.40 *	(1.95)
Fall Score	0.45 **	(0.08)
Intercept	1.81	(1.75)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Fall scores centered on fall score mean.

Appendix C: Findings from Robustness Analysis

Our robustness analysis employed a mixed effects model to estimate a pair specific effect of treatment. Like the primary analysis, this is an ANCOVA approach, modeling the Winter or Spring deviation from the mean Fall score as a function of the Minnesota Reading Corps program and the Fall mean centered scores. The model for each outcome for the *ith* student in school *j* in pair *k* is

$$y_{ijk} - \bar{x} = b_0 + b_1 MRC_{jk} + b_3\left(x_{ijk} - \bar{x}\right) + b_4 Age_{ijk} + u_k + q_k MRC_{jk} + e_{ijk},$$

where β_0 is the average growth for the average student in the comparison schools, β_1 is the average effect of the Minnesota Reading Corps program (the difference between treatment and comparison students for the average growth of the average student), β_3 is the effect of the Fall score, β_4 is the fixed age effect (when we pool the 4- and 5-year old students), u_k is the pair level random effect of the average growth for the average students at the comparison sites, q_k the pair level random effect for the Minnesota Reading Corps program effect, and e_{ijk} is the within-school student level error term. With this model, we estimate the pair specific treatment effect,

$$EFFECT_k = b_1 + q_k,$$

and test whether the mean effect is different than 0.

Exhibit C.1 below indicates that, on average, the mean of the pair effects are all different than 0. This finding indicates that the observed effects from the primary analysis are not the result of a small number of exemplary sites, but a significant effect across multiple school pairs.

Exhibit C.1: Average pair effects

			95% Confidence Interval	
	Mean Effect	SE(Mean Effect)	Low	High
Letter Sounds				
Winter	0.67	0.28	0.09	1.25
Spring	3.27	0.58	2.08	4.46
Rhyming Fluency				
Winter	1.47	0.32	0.81	2.13
Spring	3.33	0.17	2.98	3.68
Letter Names				
Winter	1.65	0.59	0.44	2.86
Spring	3.23	0.77	1.63	4.83
Picture Names				
Winter	3.27	0.40	2.45	4.10
Spring	4.18	0.33	3.50	4.86
Alliteration				
Winter	1.41	0.14	1.13	1.70
Spring	2.88	0.13	2.62	3.15

N = 25 site pair, models based on 4- and 5-year-olds

Appendix D: IES What Works Clearinghouse Analysis

The purpose of this section is to apply the IES What Works Clearinghouse (WWC) standards to the results presented for all students in this report. As this is a quasi-experimental design, the best possible rating is to meet standards with reservations.

In order to meet this standard, we must show balance on the baseline scores between program and comparison groups. Balance is expressed as a Hedges' *g* effect size that expresses the difference between the treatment (program) and comparison groups in standard deviation units. If this metric is smaller in magnitude than 0.05, the analysis can be presented without including the baseline as a covariate. If this effect size is smaller in magnitude than 0.25, but greater than 0.05, then the baseline score must be included as a covariate. When the effect size is greater than 0.25, the test cannot meet standards, even if the measure is included in the model as a covariate.

As detailed in our methodology section, we included the baseline (Fall score) as a covariate in all analyses. We then computed effect sizes for all outcomes for each age group and considered the results presented in the report. The results of this procedure are presented in Exhibit D.1 below.

The effect sizes at baseline are less than 0.25 in magnitude for all outcomes, with the exception of the Spring rhyming outcome for 3-year-old students. If the WWC standard is applied, this one significant finding in the report would be overturned due to a large difference in the Fall (baseline) score. All other findings in the report are confirmed by the WWC analysis.

Exhibit D.1: Results of WWC Balance Testing

	Report Results		Fall Score	
	Winter	Spring	Effect Size	WWC Results
3-year-olds				
Rhyming	Not sig.	Sig.	-0.38	Result overturned
Letter Names	Not sig.	Not sig.	-0.37	NA
Picture Names	Not sig.	Sig.	-0.05	Result stands
Alliteration	Not sig.	Not sig.	-0.10	NA
4- and 5-year-olds				
Letter Sounds	Not sig.	Sig.	-0.07	Result stands
Rhyming	Sig.	Sig.	-0.19	Result stands
Letter Names	Sig.	Sig.	-0.23	Result stands
Picture Names	Sig.	Sig.	-0.12	Result stands
Alliteration	Sig.	Sig.	-0.22	Result stands

Appendix E: Glossary

AmeriCorps: AmeriCorps is a national service program run by the Corporation for National and Community Service that engages members to serve at nonprofit organizations, public agencies and faith-based organizations nationwide. In exchange for their service, AmeriCorps members receive a modest living stipend and Education Award. Minnesota Reading Corps is the nation's largest state AmeriCorps program.

Benchmark: A standard score above which students are considered "on-track" for grade level achievement. Minnesota Reading Corps lists grade and season (i.e., Fall, Winter, Spring) appropriate benchmark scores for each general outcome measure (i.e., FAST and IGIDI). Students' scores on benchmark assessments determine their eligibility for Minnesota Reading Corps services and serve as baseline data to determine students' improvements as a result of the program.

"Big Five" Transitions: Big Five transitions are brief songs or games that are conducted with students during the time between scheduled activities. This time may include when students are waiting in line to put their coats on, or go outside to play, or during clean up time. The activities focus on at least one of the "Big 5" Essential Early Literacy Predictors. Examples of Big Five Transitions are the "What is it? Bag" game, a Rhyme Song, Alliteration Game, a "Letters have Names/Sounds" song.

CNCS: The Corporation for National and Community Service (CNCS) is a federal agency that engages more than 4 million Americans in service through Senior Corps, AmeriCorps, and the Social Innovation Fund, and leads President Obama's national call to service initiative, United We Serve. CNCS is the primary federal funder of the Minnesota Reading Corps program, and commissioned the current evaluation of the Minnesota Reading Corps.

Community Corps: Community Corps members are embedded in preschool classrooms and collaborate with the classroom's lead teacher to help develop children's language and emergent literacy skills in preparation for kindergarten. Community Corps members are responsible for enhancing the literacy-rich environment within the classroom, implementing Tier 2 and 3 interventions and conducting progress monitoring for students.

Educator Corps: Educator Corps members are current employees who are in a teaching position at the site, typically lead teachers or assistant teachers. This member continues to fulfill their regular teaching responsibilities, but also incorporates specific Minnesota Reading Corps strategies in their instruction.

ELLCO: The Early Language & Literacy Classroom Observation (ELLCO) tool is used to assess five key elements of a classroom's literacy environment: classroom structure, curriculum, language environment, books and book reading,

print and early writing. According to the ELLCO, a "Literacy Rich Classroom" is one that embeds literacy activities among daily routines.

Formative Assessment for Teachers (FAST): Benchmark assessments developed by the University of Minnesota. The Minnesota Reading Corps program conducts these 1-minute assessments in the Fall, Winter and Spring, which include: 1) Test of Letter Names, 2) Test of Letter Sounds, 3) Test of Nonsense Words; and 4) Curriculum-Based Measurement (CBM)-Reading.

Head Start: Head Start is a Federal early childhood program designed to promote school readiness for low-income pre-Kindergarten students by enhancing their cognitive, social and emotional development. Through Head Start programs, enrolled children and families can also receive health, nutrition and other social support programs depending on eligibility.

IGDI 1.0: The Individual Growth and Development Indicators (IGDI) is a set of standardized, individually administered assessments that are used to evaluate children's language and emergent literacy skills. IGDIs assess three key areas of emergent literacy: 1) Rhyming (Phonological Awareness); 2) Picture Naming (Vocabulary); and 3) Alliteration (Phonological Awareness).

Internal Coach: An individual trained by the Reading Corps to provide on-site literacy support and oversight to the Minnesota Reading Corps AmeriCorps member. Internal Coaches provide an on-site orientation for the member, develop a daily schedule, assist in the implementation of literacy assessments, conduct integrity checks of assessment and intervention implementation, review student data and ensure the member is accurately reporting student data. The Internal Coach is a school employee, not a Minnesota Reading Corps member or staff person.

K-3: Kindergarten through third grade

Literacy rich schedule: Members in the PreK program work to implement and support a standard instructional regime/schedule that focuses on the "Big Five" emergent literacy skills (i.e., conversation skills, vocabulary and background knowledge, book and print rules, phonological awareness-rhyming and alliteration, and alphabetic knowledge). Members assist the teaching team in implementing the literacy rich schedule and fostering a literacy rich classroom environment as defined by the Early Language and Literacy Classroom Observation (ELLCO), including name chart, theme-related books and props in five or more centers, sign-in area, writing center, word wall, etc.

Master Coach: Provides literacy coaching support to Internal Coaches and AmeriCorps members at multiple sites. The Master Coach schedules regular on-site visits to support and guide the site and its members in fulfilling the Minnesota Reading Corps program goals and ensures fidelity of implementation. The Master Coach provides training

to members, conducts integrity checks, and reviews students' progress monitoring data. The Master Coach is an experienced literacy educator who serves as a consultant to Minnesota Reading Corps.

Member: A person to who serves in the Minnesota Reading Corps AmeriCorps program. Member may refer to a volunteer in the K-3 program, or a PreK Educator Corps or Community Corps. Members deliver the one-on-one tutoring (PreK and K-3) and support implementation of the literacy rich schedule in PreK classrooms. In recognition of their service, members receive a modest living stipend and Education Award.

MDE: Minnesota Department of Education.

Minnesota Reading Corps: The Minnesota Reading Corps was started in 2003 to provide reading and literacy tutoring to children in PreK programs and students in Kindergarten through third-grade. The goal of the program is to ensure that students become successful readers and meet reading proficiency targets by the end of the third grade. Minnesota Reading Corps engages AmeriCorps members to provide literacy enrichment and tutoring services to PreK students. AmeriCorps members serve as one-on-one tutors and provide research-based interventions to both PreK and K-3 students who are just below proficiency in reading. As of the 2013-2014 school year, more than 1,100 AmeriCorps members implemented the program in 712 schools or sites and 213 school districts across the state of Minnesota.

PreK: Preschool.

Program Coordinator: An employee of Minnesota Reading Corps, responsible for providing administrative oversight to the Minnesota Reading Corps program on a regional level, including member management, site management, and compliance with all AmeriCorps regulations. The Program Coordinator oversees regional recruitment efforts, works together with service sites in the interviewing, selection, and placement process for members.

Progress monitoring: A scientifically-based practice using regular assessments to track students' academic performance and evaluate the effectiveness of an intervention. Progress monitoring data helps teaching teams determine the effectiveness of interventions then make adjustments to instruction to ensure students reach their next benchmark target. For K-3, progress monitoring is conducted with all students receiving Minnesota Reading Corps tutoring each week by members using 1-minute fluency tests. For Pre-K students receiving Tier 2 and 3 one-to-one intervention services, progress monitoring occurs monthly.

RtI: Response to Intervention (RtI) is a practice of academic and behavior interventions designed to identify and provide early effective assistance to underperforming students. Research-based interventions are implemented and frequent progress monitoring is conducted to assess student response and progress. When students do not make progress, increasingly more intense interventions are introduced.

SEEDS: A relationship-based instructional approach that maps out for teachers five ways to intentionally interact with children in order to promote academic growth and social/emotional well-being. SEEDS high quality interactions include the following five elements:

- *Sensitivity:* Look, listen, and ask questions to become aware of each child's needs, thoughts, abilities and feelings;
- *Encouragement:* Use intentional affirmations and positive non-verbal communication to create a shared positive learning environment;
- *Education:* Embed the "Big 5" literacy skills in daily routines (vocabulary, conversation, phonological awareness, book and print rules, and letter knowledge);
- *Development of Skills Through Doing:* Help children explore their world through hands-on learning; and
- *Self-Image Support:* Balance the SEEDS quality interactions to support a child's feeling of being respected and capable.

"Strive for Five": A strategy used to intentionally create extended conversations between children and adults in an interaction. It can be between one adult and one child or one adult and a small or large group of children. It was designed to encourage adults to go beyond the typical interaction that adults have with children. It encourages adults to "watch, wait, and listen", make a comment, ask a question, send a positive non-verbal message that the adult is listening, and build a conversational volley back and forth.

ServeMinnesota: State Commission on AmeriCorps programs in Minnesota and responsible in Minnesota Law for Minnesota Reading Corps.

Service hours: The required hours of service AmeriCorps members must complete in order to fulfill their 11 months of service to AmeriCorps, and in return receive a living allowance and an education award to pay for college or pay back student loans. All full-time members, K-3, Community Corps, and Educator Corps, must complete 1700 hours of service. Part-time members must complete 920 hours. Service hours can be fulfilled not only through members' time tutoring or working in the classroom, but also through participation in community and other school activities.

SMART goals: These type of goals are Specific, Measurable, Attainable, Relevant, and Timely. Within the context of the Minnesota Reading Corps PreK program, members are required to identify smart goals, and Internal Coaches are required to discuss them on a monthly basis. This process helps to ensure intervention integrity.

Summer Institute: A multi-day training program conducted over the summer to introduce new and old members, Internal Coaches, and Master Coaches to the Reading Corps program. The Institute consists of learning about the theories behind the program, the techniques used to implement the program, and the administrative components of the program. Education experts train Members and Coaches on all aspects of the program, and also provide time for practicing the techniques and the interventions. The Institute is also the time when most members will meet their Internal and Master Coaches for the first time.

Tier 1-3: Tier 1, 2, and 3 are the three "tiers" of the multi-layered instructional process at the core of the RtI model. Student scores on general outcome measures (e.g., FAST or IGDI) referenced to specific benchmarks determine a student's tier placement. The instruction that is then provided to students is based upon their respective tiers. Tier 1 students, approximately 75-80% of the population, are at the "Universal Level" and benefit from the standard whole class core literacy curriculum. They do not require supplemental instruction. Students who score in Tier 2 range, 15-20%, are those whose assessment scores are below the expected levels of achievement (benchmark) and are at risk for academic failure but are still above levels considered to indicate a high risk for failure. Tier 2 students typically are eligible for supplemental small group instruction. Students whose scores place them into Tier 3, approximately 5-10% of students, are considered to be at high risk for academic failure. They are typically offered one-to-one supplemental interventions and individualized educational plans.

- *Tier 1 Instruction:* In PreK programs, this is instruction that students receive in the general education classroom. It includes Reading Corps directed intentional teaching with embedded and explicit instruction. In K-3 programs, this is considered the core literacy instruction provided in the classrooms for all students.
- *Tier 2 Instruction:* Provides additional, more intense instruction to children identified as needing extra help in targeted skill areas. Tier 2 instruction is in addition to Tier 1 instruction.
- *Tier 3 Instruction:* Provides the most intense intervention approach for children identified as needing extra help in a targeted skill area. Tier 3 instruction builds onto Tier 2 instruction by providing more individualized and intense instruction.