

External Evaluation of the PAIRS Reading Intervention

David A. Klingbeil, PhD

Prepared for Ampact and ServeMinnesota

11/1/2023

## Contents

Background .....	3
The PAIRS Intervention .....	3
Evaluation Plan .....	3
Sampling.....	3
Inclusion Criterion .....	4
<i>Table 1. CBMreading Benchmarks by Grade and Season</i> .....	4
Analysis Plan.....	4
Creation of a Comparison Group .....	5
Impact Analysis .....	5
Exploratory Analyses.....	5
Missing Data.....	6
Results .....	6
Participants.....	6
<i>Table 2: Demographics of Students at Experimental Schools</i> .....	6
Demographics for the Analytic Sample.....	7
<i>Table 3: Demographics for the Analytic Sample by Grade</i> .....	7
<i>Table 4: Descriptive Statistics for Continuous Covariates</i> .....	8
<i>Table 5: Correlations between Continuous Variables</i> .....	8
Matching .....	8
Preliminary Data Transformation.....	8
Overall Impact .....	11
<i>Table 6: Results by Grade</i> .....	11
Exploratory Analyses.....	12
Effect of PAIRS on Spring TRF Scores .....	12
<i>Table 7: Correlations between Continuous Variables</i> .....	12
<i>Table 8: Effect of PAIRS on Spring TRF Scores by Grade</i> .....	12
Summary .....	13
References.....	14

## Background

Reading Corps provides supplemental reading interventions to students in kindergarten through 3rd grade in several states (Markovitz et al., 2022). Historically, Reading Corps interventions to second and third grade students in a 1-on-1 format where an AmeriCorps member (i.e., tutor) conducts each intervention session with one student. Interventions are scheduled to occur 5 days per week for 20 min per session. Tutors work with approximately 15 students, across kindergarten through third grade, per day. As such, this is a relatively intensive level of support.

The purpose of this external evaluation was to examine whether the effect of a paired-version of a reading fluency intervention ("PAIRS") was not unacceptably worse than the effect for the traditional 1-on-1 fluency interventions used by Reading Corps. In the PAIRS intervention, a Reading Corps tutor conducts a reading fluency intervention with two students at a time. Due to the similar number of opportunities to respond, Reading Corps hypothesized that the PAIRS intervention would have similar effects to the 1-on-1 reading fluency interventions (i.e., business-as-usual). If so, the PAIRS intervention would provide a way to increase the number of students who receive Reading Corps tutoring services and increase the efficiency of the support Reading Corps tutors provide to their partnering school sites.

## The PAIRS Intervention

Given the pressing need to increase the number of students who receive tiered intervention, ServeMinnesota and Ampact partnered to create an intervention that targeted oral reading fluency and could be administered in a paired format.

During the intervention, students read a connected text passage that both students can read with at least 95% accuracy. Prior to starting the intervention, the tutor breaks up the passage into two or four sentence sections. The students read through each section of the passage four different ways. First, the students alternate reading sentences. Second, the students alternate reading sentences but change which student reads first. Third, the tutor reads the passage aloud with the two students who are asked to match the tutor's expression and pace. Fourth, the students read the passage together without the tutor. Similar to the traditional text reading fluency interventions used in Reading Corps, the tutor provides immediate error correction and specific praise to the students throughout the intervention.

## Evaluation Plan

### Sampling

Ampact and ServeMinnesota staff selected a sample of 95 schools in Minnesota and Wisconsin based on the Reading Corps coaching specialist assigned to the school. All schools in the sample were supported by 1 of 5 coaching specialists. Therefore, schools in the control group had the same coaching specialist.

ServeMinnesota staff randomly assigned schools in the sample to one of two conditions: (1) piloting the PAIRS intervention or (2) serving as the business-as-usual control group. Not all of the schools randomly assigned to the experimental condition ( $j = 49$ ) or control condition ( $j = 46$ ) had a tutor placed during the August or October start date. No students were receiving tutoring at those sites prior to November 1<sup>st</sup> and those schools were all excluded from the analysis. In the final analytic sample, there were 27 schools randomly assigned to the PAIRS condition and 37 schools assigned to the control condition.

## Inclusion Criterion

To participate in the study, students had to meet the following criteria: (1) be enrolled in Grades 2 or 3, (2) meet Reading Corps program eligibility, (3) be assigned to a reading fluency intervention, and (4) be added to a Reading Corps tutor's caseload before November 1<sup>st</sup>. This last criterion was to ensure that only students who were added after fall benchmarking were included in the analysis.

The same program eligibility requirements were used across conditions in the study. Reading Corps tutors collect benchmarking data using the FastBridge curriculum-based measures of reading (CBMreading; Christ et al., 2015) during a 2-week period in September or October. To be eligible for the program, students' text reading fluency (TRF) score (or the number of words read correct per minute) had to be below the FastBridge CBMreading benchmarks (Table 1) for their grade level.

*Table 1. CBMreading Benchmarks by Grade and Season*

Grade	Fall		Winter		Spring	
	Words Read Correct / Min	National Percentile	Words Read Correct / Min	National Percentile	Words Read Correct / Min	National Percentile
2	63	45 <sup>th</sup>	97	51 <sup>st</sup>	116	54 <sup>th</sup>
3	100	51 <sup>st</sup>	122	49 <sup>th</sup>	135	46 <sup>th</sup>

*Note.* National percentiles based on the vendor-provided technical documentation (Christ et al., 2015).

## Analysis Plan

Schools in the experimental condition were able to select students to participate in PAIRS but were not required to enroll all eligible students in the PAIRS intervention. ServeMinnesota requested that schools in the experimental condition purposefully assign a minimum of four students to receive PAIRS after fall benchmarking. Therefore, it did not seem reasonable to assume that the students at the treatment schools, who were not assigned to PAIRS for logistic or other reasons, would provide a suitable comparison group. This precluded the use of within-cluster pair matching. Although random assignment occurred at the cluster level, all students in the randomly assigned sites were NOT assigned to the intervention. Therefore, the randomization at the cluster level was presumed to be incidental to the actual study comparison. Thus, it was not reasonable to consider this a cluster randomized controlled trial. Given these two caveats, I estimated the average treatment effect on the treated (ATT). The ATT estimates the treatment effect of PAIRS for the population of Reading Corps participants who would be most likely to receive PAIRS, rather than all students eligible for Reading Corps (Austin, 2011). It is unlikely that all students would be a good fit for PAIRS or that all students could be paired within a site due to logistical factors (e.g., scheduling).

I had originally planned to conduct the matching and impact analysis using the aggregated sample of second and third grade students (and including grade as a predictor in the matching and analytic models). This was the approach used to analyze results from a recent randomized controlled trial of Reading Corps (Markovitz et al., 2022). However, after fitting a preliminary match on the aggregated sample, I investigated the pairs (see below) and found that students were matched across grades. This did not seem to be a suitable comparison for this impact analysis. Therefore, I stratified the sample by grade prior to conducting the matching and subsequent impact analyses and did not use grade as a covariate.

### Creation of a Comparison Group

I used matching to select a suitable comparison group from students at the control schools that received typical 1-on-1 Reading Corps services. The purpose of matching is to refine the analytic sample so that the non-random treatment assignment is essentially independent from the measured covariates in the final analytic sample (Greifer & Stuart, 2021). I originally planned to use nearest neighbor matching without replacement to create a comparison group (1:1 ratio). However, more recent guidance suggests that analysts try multiple types of matching to evaluate the covariate balance resulting from each, before conducting the actual analysis of interest (Greifer & Stuart, 2021). Therefore, I assessed covariate balance after conducting the match with three different methods, nearest neighbor matching with replacement (1:1 ratio), optimal full pair matching, and coarsened exact matching.

To do this, I fit a model that estimated the probability of receiving PAIRS based on students' Sex, race/ethnicity, EL status, and fall TRF scores. After conducting the matches, I investigated the balance in these covariates as well as the between-groups balance in intervention dosage. Please note, however, that I did not use dosage in the creation of the propensity scores as it could be influenced by the treatment itself (i.e., an endogenous covariate; What Works Clearinghouse, 2023).

### Impact Analysis

After creating the matched sample, I used linear regression to estimate the average treatment effect of receiving PAIRS. In the final analytic model, I predicted winter TRF scores by PAIRS assignment, controlling for fall TRF and intervention dosage as covariates. PAIRS assignment was entered as a binary treatment indicator (1:1 intervention = 0, PAIRS = 1).

I used the `marginalEffects` R package to estimate the effect of the intervention on the treatment group (ATT) using `g-computation` (Wang et al., 2017). I used cluster-robust standard errors to account for dependence between the matched pairs (Greifer, 2023) and bias-corrected accelerated bootstrap confidence intervals to estimate the 95% CI around the estimated effect of receiving PAIRS.

Finally, I used the marginal means to estimate the standardized mean difference adjusted for small sample bias (i.e., Hedges'  $g$ ) following the guidance in the What Works Clearinghouse handbook. The mean difference was estimated after adjusting for fall TRF scores and intervention minutes. The pooled standard deviation was estimated using the unadjusted within-group standard deviations estimated from the pair matched sample.

### Exploratory Analyses

I also conducted an exploratory analysis of whether there were any differences associated with receiving PAIRS in students' Spring TRF scores, after controlling for fall TRF and intervention dosage. Intervention dosage was estimated across the entire school year instead of between fall and winter benchmarking.

I also planned to conduct an exploratory analysis to investigate whether student demographics were associated with differential effects of the PAIRS intervention. However, these exploratory analyses were not conducted due to the small sample sizes in demographic groups. Prior evidence indicates that Reading Corps interventions are similarly effective across demographic groups (Markovitz et al., 2018) but future evaluations could ensure a more equal distribution of students across demographic groups of interest (e.g., English learners) to determine whether the findings of this analysis generalize across student subgroups.

## Missing Data

After creating the analytic sample, I investigated missingness on the dependent variable and covariates to be used in the matching process as well as the analytic models. These variables included intervention minutes, race/ethnicity, Sex, English learner status, fall TRF scores, and winter TRF scores. No students were missing information regarding intervention dosage. A total of 22 students were missing data regarding their race/ethnicity and one student was missing data regarding their Sex. Regarding English learner status, 37 students were listed as unknown. For each of these three variables, I coded missing data or unknown status as Not Reported.

There was no missing data in the treatment group (i.e., received PAIRS) on the fall TRF scores. There were 45 students in the control schools who were missing the fall TRF scores ( $n = 31$  in Grade 2,  $n = 14$  in Grade 3). Missingness was most pronounced on the winter TRF scores, which was the dependent variable of interest. There were 83 students missing winter TRF scores, including 17 students who received PAIRS ( $n = 12$  in Grade 2,  $n = 5$  in Grade 3).

## Results

### Participants

Demographics for the students at the experimental schools are shown in Table 2. I conducted a series of chi-squared tests to evaluate whether student demographic variables were associated with assignment to the PAIRS intervention. Student grade, race/ethnicity, English learner status, or Sex were not associated with assignment to PAIRS. Please note that the “Not assigned to PAIRS” students were not actually included in the analytic sample.

*Table 2: Demographics of Students at Experimental Schools*

	Assigned to PAIRS		Not Assigned to PAIRS		$\chi^2$	df	p
	n	%	n	%			
Grade					2.53	1	.112
2 <sup>nd</sup>	71	31.3%	156	68.7%			
3 <sup>rd</sup>	82	38.5%	131	61.5%			
Race or Ethnicity					4.37	6	.626
American Indian or Alaska Native	1	0.6%	4	1.4%			
Asian	5	3.3%	5	1.7%			
Black	10	6.5%	31	10.8%			
Hispanic or Latinx	9	5.9%	13	4.5%			
Two or More	9	5.9%	15	5.2%			
Native Hawaiian or Pacific Islander	0	0%	0	0%			
Not Reported	23	15.0%	36	12.5%			
White	96	62.8%	183	63.8%			
English Learner					3.24	2	.198
Yes	19	12.4%	31	10.8%			
No	111	72.5%	228	79.4%			
Not Reported	23	15.0%	28	9.8%			
Sex					4.82	2	.090
Female	88	57.5%	141	49.1%			
Male	64	41.8%	137	47.7%			
Not Reported	1	0.01%	9	3.1%			

*Fall Text Reading Fluency Benchmarking*

	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>
Words Read Correct Min (Grade 2)	41.3	12.3	27.4	18.7	6.58	196	< .001
Words Read Correct Min (Grade 3)	68.9	19.3	64.2	21.5	1.66	186	.098

*Note.* The students in the Not Assigned to PAIRS column were excluded from all further analyses.

I also evaluated whether students’ fall benchmarking scores were associated with PAIRS assignment. I conducted these analyses separate by grade as students are benchmarked with grade-level materials. In Grade 2, students who received PAIRS had significantly higher words read correct per minute during fall benchmarking than students who did not receive PAIRS. This was not true in Grade 3, students who received PAIRS did not have significantly higher words read correct per minute during fall benchmarking (see Table 2). Together, these results suggest that school sites may have assigned higher performing students in Grade 2 to PAIRS but that a number of unmeasured factors may have influenced PAIRS assignment.

### Demographics for the Analytic Sample

Demographics for the students at the experimental schools are shown in Table 3. I conducted a series of chi-squared tests to evaluate whether student demographic variables were associated with assignment to the PAIRS intervention. Student grade, race/ethnicity, English learner status, or Sex were not associated with assignment to PAIRS.

*Table 3: Demographics for the Analytic Sample by Grade*

	Grade 2				Grade 3			
	PAIRS ( <i>n</i> = 71)		Control ( <i>n</i> = 258)		PAIRS ( <i>n</i> = 82)		Control ( <i>n</i> = 212)	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Race or Ethnicity								
American Indian or Alaska Native	0	0.0%	6	2.3%	1	1.2%	3	1.4%
Asian	3	4.2%	6	2.3%	2	2.4%	8	3.8%
Black	7	9.9%	25	9.7%	3	3.7%	24	11.3%
Hispanic or Latinx	5	7.0%	16	6.2%	4	4.9%	11	5.2%
Two or More	7	9.9%	12	4.7%	5	6.1%	16	7.5%
Native Hawaiian or Pacific Islander	0	0.0%	0	0.0%	0	0.0%	1	0.5%
Not Reported	3	4.2%	8	3.1%	12	14.6%	15	7.0%
White	46	64.8%	185	71.7%	55	67.1%	135	63.4%
English Learner								
Yes	8	11.3%	29	11.2%	11	13.4%	30	14.2%
No	55	77.5%	213	82.6%	62	75.6%	166	78.3%
Not Reported	8	11.3%	16	6.2%	9	11.0%	16	7.5%
Sex								
Female	39	54.9%	140	54.3%	49	59.8%	105	49.5%
Male	32	45.1%	118	45.7%	32	39.0%	107	50.5%
Not Reported	0	0.0%	0	0.0%	1	1.2%	0	0.0%

Next, I compared the PAIRS and Control groups on their fall TRF scores along with their text reading accuracy. Results are shown in Table 4. Second grade students assigned to PAIRS had statistically higher fall TRF scores than second grade students assigned to typical Reading Corps interventions in the control schools,  $t(172) = 6.71, p < .001$ . In addition, the between group differences in text reading accuracy was

statistically significant in Grade 2,  $t(259) = 5.65, p < .001$ . Together with the results evaluating differences within the experimental schools, it seems that school sites assigned second grade students to PAIRS who were reading with higher fluency and accuracy than the average 2<sup>nd</sup> grade student who received the typical 1-on-1 Reading Corps intervention.

In Grade 3, the observed differences in fall TRF scores were not statistically significant  $t(156) = 1.28, p = .204$ . Neither were the observed differences in fall text reading accuracy,  $t(196) = 1.49, p = .139$ . In combination with the results evaluating differences within the experimental schools, it seems that the third-grade students assigned to PAIRS were reading with similar text reading fluency and accuracy at fall benchmarking as the average third student who received the typical 1-on-1 Reading Corps intervention.

*Table 4: Descriptive Statistics for Continuous Covariates*

	Grade 2				Grade 3			
	PAIRS (n = 71)		Control (n = 258)		PAIRS (n = 82)		Control (n = 212)	
	M	SD	M	SD	M	SD	M	SD
Fall Benchmarking								
WRCM	41.3	12.3	29.6	18.1	68.9	19.3	65.7	19.3
Text Reading Accuracy	86.2%	9.6%	76.0%	21.1%	94.1%	5.4%	92.9%	7.0%
Intervention Dosage								
Sessions	46.2	13.3	44.9	15.6	45.6	15.5	44.9	15.6
Total Minutes	896	264	946	290	898	310	878	312
PAIRS Minutes	742	338	N/A	N/A	808	344	N/A	N/A

There were no significant differences between the number of intervention minutes between students assigned to PAIRS or the typical 1-on-1 intervention. On average, second grade students received 83% of their intervention minutes in the PAIRS intervention (range = 6% to 100%). Third grade students assigned to PAIRS received 90% of their instructional minutes in PAIRS (range = 12% to 100%).

*Table 5: Correlations between Continuous Variables*

Variable	1.	2.	3.	4.
1. Fall WRCM	--	.75*	-.16*	.84*
2. Fall Text Reading Accuracy	.66*	--	-.09*	.73*
3. Intervention Minutes	-.13	-.08	--	-.21*
4. Winter WRCM	.77*	.59*	-.26*	--

Note. Grade 2 values are shown above the diagonal. Grade 3 values are shown below the diagonal. \*  $p < .0125$

## Matching

### Preliminary Data Transformation

Prior to conducting the match process, I collapsed Asian and Native Hawaiian or Pacific Islander categories due to small sample sizes. I also transformed some of the continuous variables to facilitate interpretation of the model intercepts. First, I transformed intervention minutes into intervention hours. Next, I centered students' fall TRF scores and intervention hours using the grade-level average values. Finally, I dropped the one student who was missing Sex.



I created a matched sample using nearest neighbor without replacement (1:1 ratio), optimal full matching, and coarsened exact matching. Figures 1 and 2 show the covariate balance, for Grades 2 and 3 respectively, in terms of mean differences between the matched treatment and control groups. The covariate balance in the unadjusted sample is also shown. I used the threshold of  $> .10$  or  $< -.10$  as indicative of remaining imbalance in the sample (Griefer & Stuart, 2021).

The unadjusted sample (pre-match) was relatively balanced on the race/ethnicity variables except students who identified as multiracial. However, there was substantial differences in the mean centered TRF scores as well as the eventual intervention dosage received. All three matching methods reduced the imbalance to tolerable levels on each of the covariates. The optimal full matching approach further reduced imbalance in fall TRF scores, but it increased the imbalance in the sample in terms of the proportion of White students (Grade 3). The full sample had a higher effective sample size than the 1:1 nearest neighbor with replacement in grade 2, but it resulted in a smaller effective sample size in Grade 3. Therefore, I used the analytic sample created with the nearest neighbor method without replacement to investigate the overall impact of PAIRS in both grades.

Figure 1: Covariate Balance across Unadjusted and Matched Samples in Grade 2

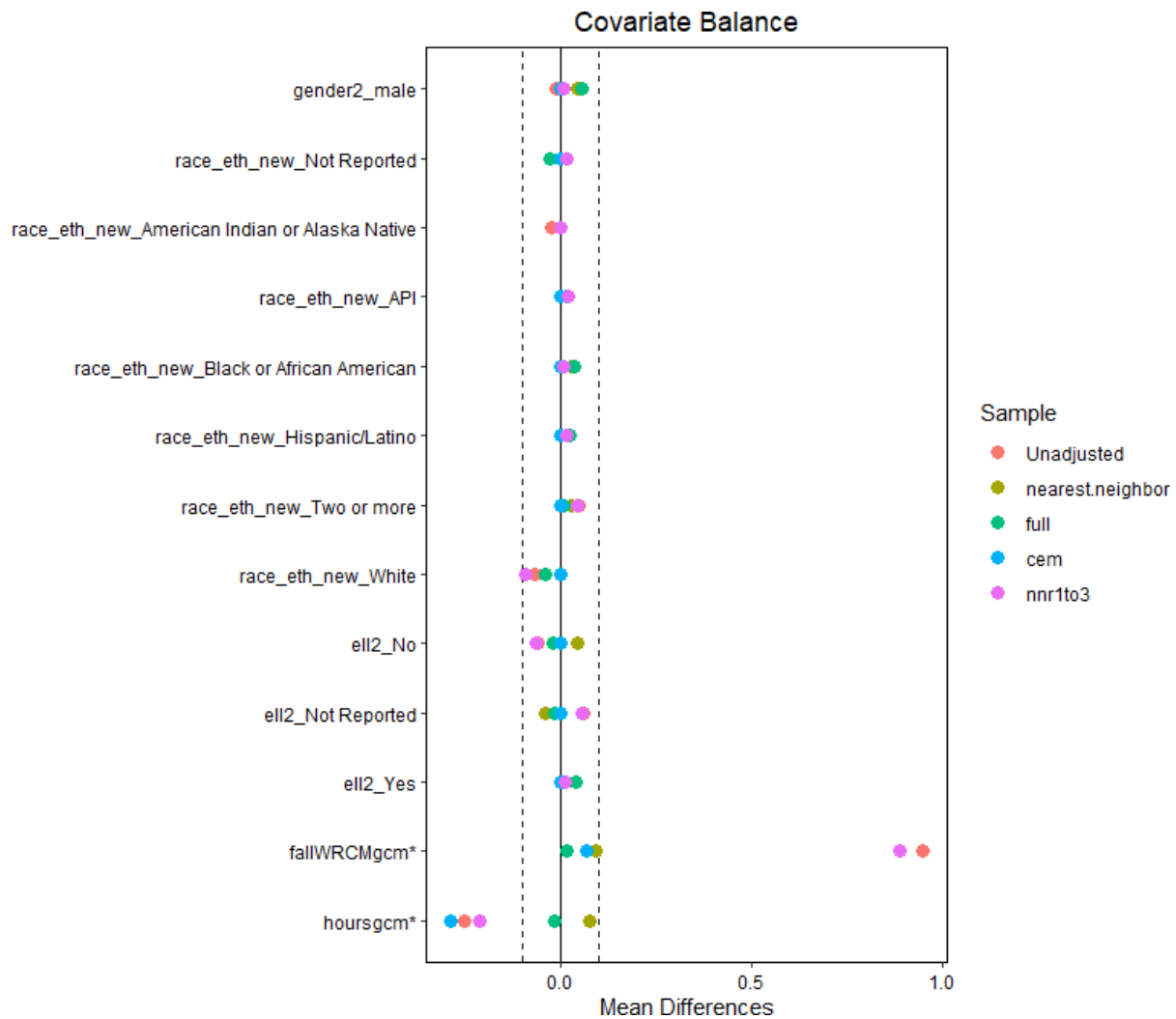
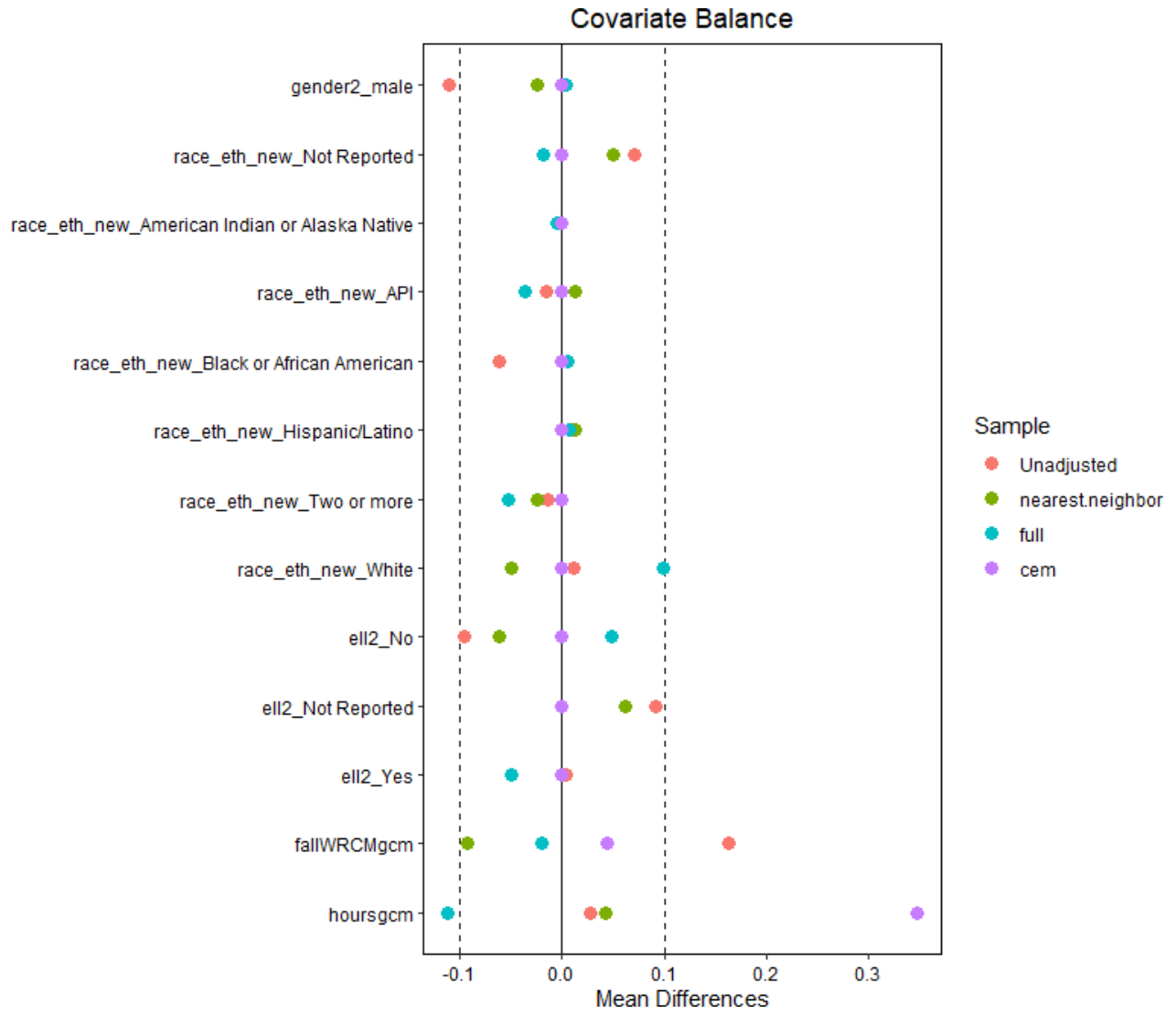


Figure 2: Covariate Balance across Unadjusted and Matched Samples in Grade 3



## Overall Impact

I fit a linear regression model with students' winter TRF scores predicted by fall TRF scores, intervention hours, and the binary indicator variable for whether the student received PAIRS (1 = Yes, 0 = No). After fitting the model, I used non-parametric bootstrapping to estimate the confidence interval for the estimated effect of being assigned to PAIRS. Results are shown in Table 6. Because the covariates were mean centered, the intercept represents the average winter TRF score for a student in the control group.

The estimated effect of receiving the PAIRS intervention, after controlling for fall TRF scores and intervention dosage, was 0.82 words read correct per minute during winter benchmarking in Grade 2 (bca adjusted 95% CI = -4.96 to 6.64). The effect of participating in the PAIRS intervention was not statistically significant.

A similar pattern was found in Grade 3. There was a relatively small, negative effect associated with receiving the PAIRS intervention of -0.47 words read correct per minute (bootstrap adjusted 95% CI = -5.12 to 3.95). However, this difference was not statistically significant.

*Table 6: Results by Grade*

Variable	Estimate	SE	t	p
Grade 2				
Intercept	63.76	2.30	27.71	< .001
Fall TRF – WRCM	1.23	0.11	10.82	< .001
Intervention Hours	-0.32	0.40	-0.81	.425
Had PAIRS	0.82	3.00	0.27	.785
Grade 3				
Intercept	100.50	1.83	54.79	< .001
Fall TRF - WRCM	0.82	0.07	11.04	< .001
Intervention Hours	-0.71	0.24	-3.03	.003
Had PAIRS	-0.47	2.27	-0.21	.837

*Note.* The fall words read correct per minute and intervention hours were centered at the mean within each grade level to facilitate interpretation of the intercept and the effect of PAIRS.

I also calculated the standardized mean difference adjusted for small sample bias (i.e., Hedges'  $g$ ) associated with receiving PAIRS (on winter TRF scores). Effect sizes were estimated using the adjusted model mean averaged across fall TRF scores and intervention hours (What Works Clearinghouse, 2023). The estimated model mean were divided by the pooled, unadjusted standard deviation in winter TRF scores in the matched sample. The small sample bias correction was estimated using the effective sample size in the optimal full matched sample. The effect sizes associated with participating in PAIRS were negligible in Grade 2 ( $g = .04$ ) and Grade 3 ( $g = -0.05$ ).

The results of this impact analysis indicate that being assigned to PAIRS had a small, non-significant effect on winter TRF scores in comparison to the typical 1-on-1 intervention services provided by Reading Corps. This point is important as these results do not mean that PAIRS did not have an effect in comparison to not receiving any intervention.

## Exploratory Analyses

### Effect of PAIRS on Spring TRF Scores

For the exploratory analysis on spring TRF scores, there were 75 students in the analytic sample who were missing the spring TRF score. I present the results of the correlations between the covariates and the spring TRF scores in Table 7. The intervention minutes variable now represents the full dosage received for the entire year, rather than the dosage between fall and winter benchmarking.

*Table 7: Correlations between Continuous Variables*

Variable	1.	2.	3.	4.
1. Fall WRCM	--	.71*	-.27*	.73*
2. Fall Text Reading Accuracy	.66*	--	-.21*	.69*
3. Intervention Minutes All Year	-.23	-.17*	--	-.31*
4. Spring WRCM	.69*	.55*	-.33*	--

*Note.* Grade 2 values are shown above the diagonal. Grade 3 values are shown below the diagonal. \*  $p < .0125$

Next, I fit a linear regression model to evaluate the overall impact of the PAIRS intervention on students' Spring TRF scores. This analysis was conducted separately by grade. Students' winter TRF scores were predicted by fall TRF scores, intervention dosage across the entire school year, and a binary indicator variable for whether the student received PAIRS (1 = Yes, 0 = No). As in the overall impact analysis, I transformed minutes into hours. I also centered the covariates at the mean for each grade level. In each model in Table 8, the intercept represents the average spring TRF score for a student in the control group at the grade-level average fall WRCM and intervention hours.

*Table 8: Effect of PAIRS on Spring TRF Scores by Grade*

Variable	Estimate	SE	t	p
Grade 2				
Intercept	88.47	2.83	31.24	< .001
Fall WRCM	0.93	0.15	6.14	< .001
Intervention Hours All Year	-0.24	0.17	-1.46	.146
Had PAIRS	-0.73	3.29	-0.22	.824
Grade 3				
Intercept	114.68	2.13	53.84	< .001
Fall WRCM	0.72	0.09	7.97	< .001
Intervention Hours All Year	-0.42	0.13	-3.41	< .001
Had PAIRS	2.18	2.54	.857	.392

*Note.* The fall WRCM and Intervention Hours were centered at the mean within each grade level to facilitate interpretation of the intercept and the effect of PAIRS.

The estimated effect of receiving the PAIRS intervention, after controlling for fall TRF scores and intervention dosage across the entire year, was -0.73 words read correct per minute during spring benchmarking in Grade 2 (95% CI = -7.27 to 5.40). In grade 3, there was a positive effect associated with receiving the PAIRS intervention on students' spring TRF score of 0.74 words read correct per minute (95% CI = -4.87 to 4.04). The effects associated with participating in PAIRS was not significant in either grade.

Although the t-test associated with the treatment indicator (Had Pairs = Yes) were not significant, these analyses may be underpowered due to sample sizes. Therefore, I still calculated the standardized mean difference adjusted for small sample bias (i.e., Hedges'  $g$ ) associated with receiving PAIRS in winter TRF scores. Effect sizes were estimated using the adjusted model means controlling for the fall TRF scores and intervention minutes (What Works Clearinghouse, 2023). The estimated model means were divided by the pooled, unadjusted standard deviation in winter TRF scores in the matched sample. The effect sizes associated with participating in PAIRS was  $g = -0.04$  in Grade 2 and  $g = 0.03$  in Grade 3.

## Summary

Results of this evaluation suggest that students assigned to the PAIRS intervention are, on average, likely to perform at a similar rate as students assigned to the typical Reading Corps interventions delivered in a 1-on-1 format. It is important to remember that the PAIRS group in second grade seemed to be higher performing than students who were not assigned to PAIRS within the treatment schools. Therefore, it is difficult to say whether PAIRS would result in similar effects for second grade students who were performing further from the fall benchmark.

It is also important to remember that school sites in the study were not asked to randomly assign students to PAIRS or use PAIRS with all students in grades 2 or 3. Sites were able to select students who they felt were a "good fit" based on student characteristics (reading or otherwise) and logistical factors (e.g., scheduling). Ampact and ServeMinnesota should consider giving the same guidance regarding PAIRS assignment to schools that may consider using PAIRS in the future.

## References

- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399-424.  
<https://doi.org/10.1080/00273171.2011.568786>
- Griefer, N. (2023). *Estimating effects after weighting*.  
<https://ngriefer.github.io/WeightIt/articles/estimating-effects.html#references>
- Griefer, N., & Stuart, E. A. (2021). Matching methods for confounder adjustment: An addition to the epidemiologist's toolbox, *Epidemiologic Reviews, 43*(1), 118-129.  
<https://doi.org/10.1093/epirev/mxab003>
- Ho, D., Imai, K., King, G., & Stuart, E. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42*(8), 1-28. <https://doi.org/10.18637/jss.v042.i08>
- Markovitz, C. E., Hernandez, M. W., Hedberg, E. C., Whitmore, H. H., & Satorius, J. L. (2018). Impact evaluation of the Minnesota reading corps K-3 program (2017-2018). University of Chicago.  
[https://americorps.gov/sites/default/files/evidenceexchange/Reading\\_Math\\_19AC208813\\_Report\\_Revised\\_508\\_1.pdf](https://americorps.gov/sites/default/files/evidenceexchange/Reading_Math_19AC208813_Report_Revised_508_1.pdf)
- Wang, A., Nianogo, R. A., & Arah, O. A. (2017). G-computation of average treatment effects on the treated and the untreated. *BMC Medical Research Methodology, 17*, 3.  
<https://doi.org/10.1186/s12874-016-0282-4>
- What Works Clearinghouse (2023). *Procedures and standards handbook* (version 5.0).  
<https://ies.ed.gov/ncee/wwc/handbooks>