

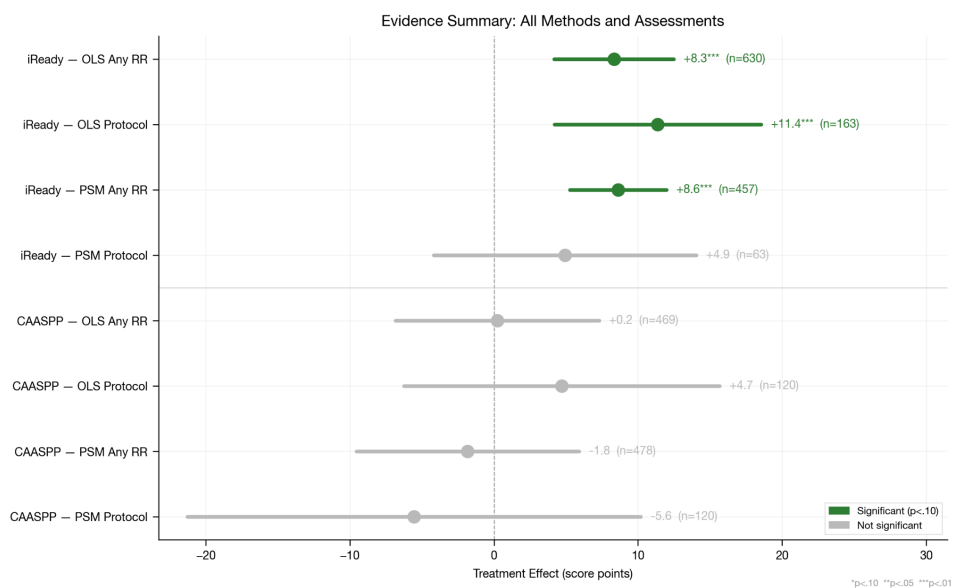
Effect of Rally Reader on Student Reading

Reading Assessment Effects in a California School District 2024-25

Abstract

Rally Reader usage at protocol (at or above the grade-level 75th percentile of usage hours) has a significant positive effect on student reading achievement, consistent with ESSA Tier 3 (Promising Evidence) criteria. This finding is supported by multiple controlled analytical methods across two independent standardized assessments.

Difference-in-difference (DiD) analyses conducted via ordinary least-squares regression (OLS) and propensity score matching (PSM) across iReady and CAASPP assessments indicate that students using Rally Reader show meaningful improvement in reading achievement relative to non-users. All four iReady analyses — the primary and more finely-measured assessment panel — show positive treatment effects, with three reaching statistical significance. CAASPP results are directionally positive for protocol-adherent students at the grade level, with Grade 5 reaching significance. No analysis produced a statistically significant negative result.



The effect is strongest for protocol-adherent students in Grade 3, who showed 132% greater growth than their control peers on iReady, with positive signals in other grades as well. On the CAASPP — California’s state summative assessment — Grade 5 protocol students scored +21.6 points higher than control peers (p=0.048). These findings reflect a single-district quasi-experimental design; multi-district replication would further strengthen the evidence base.

iReady: Treatment Effect by Grade
Protocol (P75+) vs Control | % Extra Growth = DiD / control mean gain

Grade	n (Prot / Ctrl)	DiD Effect	% Extra Growth	p-value	% Oral
Grade 2	24 / 218	+11.0	+46%	0.229	85%
Grade 3	55 / 210	+16.2*	+132%	0.086	66%
Grade 4	42 / 114	+15.0	+79%	0.203	50%
Grade 5	31 / 193	+20.0***	+479%	0.010	47%

Introduction

Rally Reader is an eReader app that combines popular books with a digital reading coach to provide real-time feedback to assist students in learning pronunciation and vocabulary, and provides reading motivation. The application offers a digital reader interface for the student's book choice. While using the application, the student may read aloud and the application will track the reader's progress along the page, assisting when they may get stuck and providing visual indicators of their progress. Alternatively, the student may read silently, akin to a more traditional e-book interface. The purpose of this application is to measure accurate diagnostics of a student's reading progress while providing them with a wide selection of reading options, assistance when they are stuck, akin to the way a parent or teacher would when reading side by side, and motivation during reading sessions, and encouragement in their post-session performance.

Background

Rally Reader was introduced to the school district at the end of the 2023-2024 school year in a pilot program before rolling out more broadly across the district in the 2024-2025 school year. With no pre-planned randomized control trial outlined ahead of the process, a post hoc analysis was conducted to draw any conclusions as to the causal effects of Rally Reader on student reading assessments for students who used Rally Reader during the '24-25 school year.

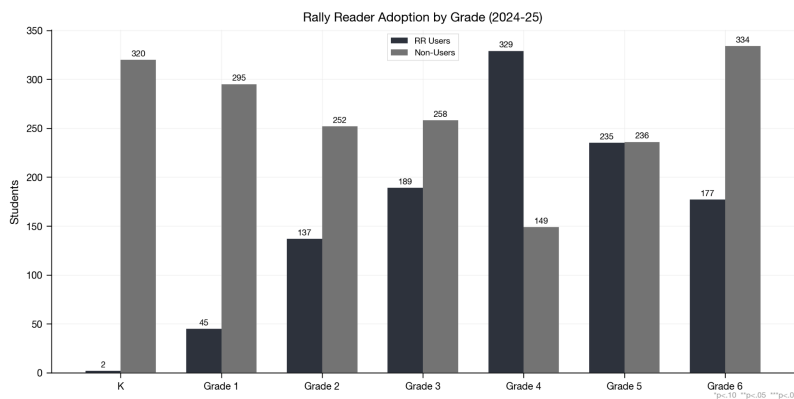
The relative randomization and breadth of Rally Reader adoption provides a reasonable dataset to draw conclusions on the efficacy of Rally Reader usage on student reading achievement. This analysis reviews two sets of standardized tests as measures of student reading performance. The primary measure is the state-administered English Language Assessment (CAASPP ELA) which is administered annually in the second half of the school year in the Spring. The second is [iReady](#), an adaptive diagnostic examination that is administered multiple times throughout the school year and offers a more precise measure of performance as compared to the blunt state standardized tests.

District Context

The school district serves approximately 5,400 elementary students (grades K-6) across seven elementary schools. Rally Reader was introduced as a supplemental reading tool in May 2024 for use on the Apple iPad, with adoption expanding across classrooms throughout the 2024-2025 school year. Adoption was teacher-driven and voluntary; no classrooms were randomly assigned to use or not use the product.

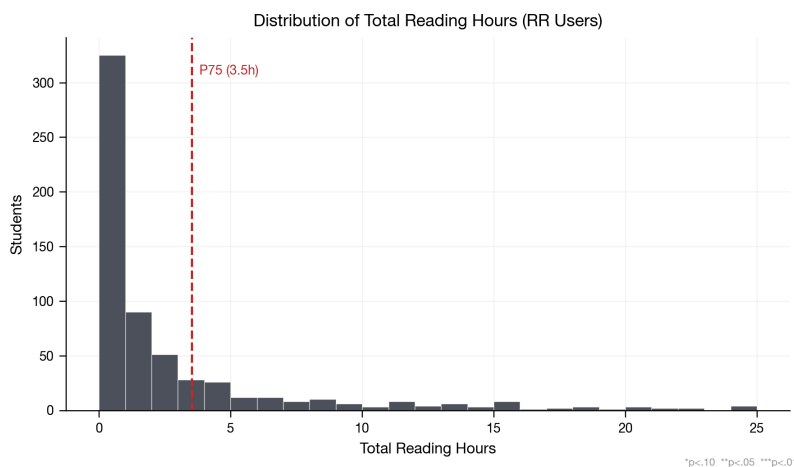
By March 2026, 1,501 students had used Rally Reader at least once. Adoption was concentrated at six core elementary schools: Chaparral (258 users), Sycamore (229), Condit (178), Sumner (167), Vista del Valle (127), and Mountain View (97). Adoption was not uniform — it rolled out in waves, with 227 students starting in May 2024 (pilot), a second wave in September-December 2024, and continued onboarding through the 2025-2026 school year.

Rally reader adoption also varied by grade, with adoption highest in grades 3-5. This imbalance is controlled for in all regression measurements and affects the outcome of DiD effects due to the expected scale-score improvement of students changing by grade, and is the reason per-grade results are reported alongside pooled estimates.



Usage Patterns

Among students who used Rally Reader during the '24-25 school year the median total usage was 1.0 hours (average 4.5 hours). The 75th percentile was 3.5 hours and the 90th percentile was 12.1 hours.



The protocol threshold in this study is defined as the grade-level 75th percentile (P75) of total usage hours among students with any Rally Reader usage. This approach accounts for differences in typical usage levels across grades and identifies students whose engagement is meaningfully above their grade peers. The P75 thresholds by grade are: Grade 2 (0.9h), Grade 3 (3.5h), Grade 4 (8.3h), Grade 5 (2.9h), Grade 6 (9.7h). Students at or above their grade's P75 threshold are classified as protocol-adherent.

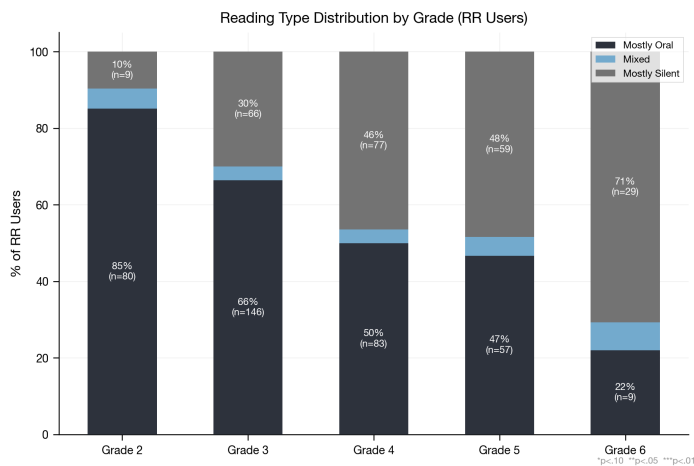
Additionally, a minimum usage floor of 2 minutes is applied to distinguish genuine engagement from incidental app-opens. Students with less than 2 minutes of total usage are reclassified as non-users for the binary "Any RR" analysis.

Reading Type

Rally reader supports two modes of reading: oral (read-aloud validated by the app) and silent. Among users in the iReady panel:

- 366 students (58%) were classified as mostly oral
- 28 students (4%) as mixed
- 240 students (37%) as mostly silent

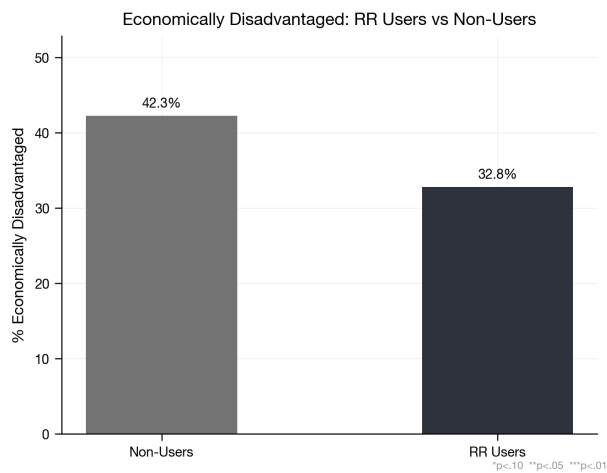
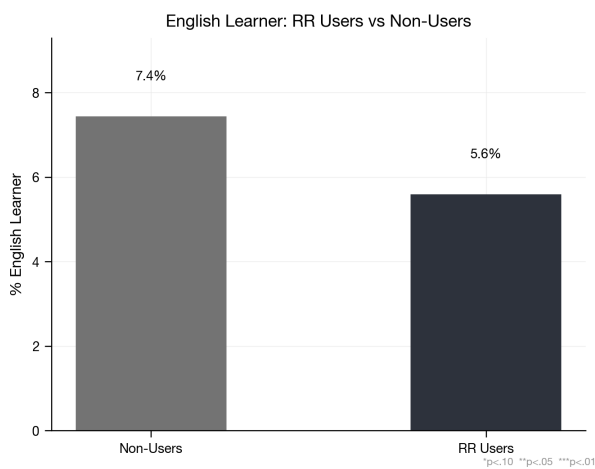
Oral reading use decreased with age: 85% of 2nd graders primarily read aloud, compared to 22% of 6th graders. Mostly silent readers logged more total hours on average (9.6h) than mostly oral readers (1.2h).

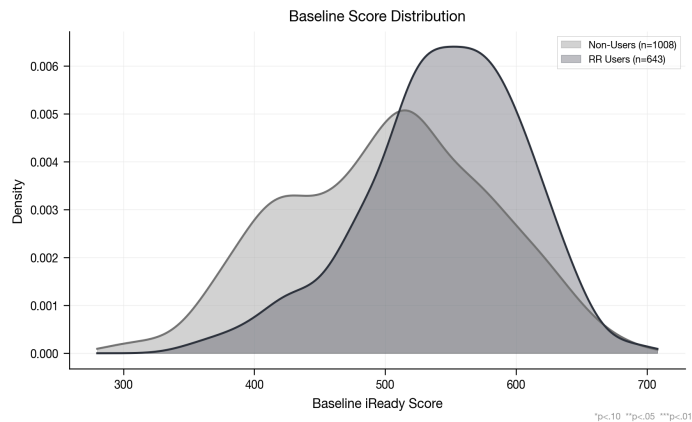
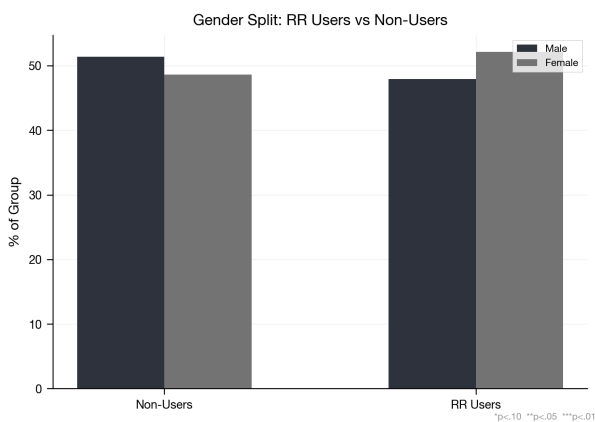


Demographics

The sample of students for iReady (grades 2-6) includes 630 Rally Reader users and 1,021 non-users. The groups are broadly comparable but not identical.

Rally Reader users are less likely to be economically disadvantaged (32.4% vs 42.4%), less likely to be English language learners (5.7% vs 7.3%), skew more towards female students (52.2% vs 48.6%), and have higher baseline scores (545 vs 503). The difference in baseline scores suggests that the tool may have been directed towards students who could benefit from additional reading practice in addition to students who are self-selecting into a reading tool that offers benefits like a broad library of reading material.





These differences are controlled for in all regression analyses. Propensity score matching provides an additional check by explicitly balancing these characteristics before estimating treatment effects.

Methodology

This study uses a quasi-experimental difference-in-differences (DiD) design to estimate the causal effect of Rally Reader on student reading achievement. Outcomes show the score improvement of treatment groups in standardized tests over the school year relative to control. Because adoption was not randomized, we compare outcomes between students who used Rally Reader and those who did not, controlling for pre-existing differences.

The study population includes students in grades 2-6 across the district. Of these, 1,501 students used Rally Reader at any point during the study period.

Assessments

Students achievement is measured using two independent standardized assessments:

CAASPP ELA – California's annual summative assessment, administered each Spring. The CAASPP panel (n=1,312) includes students with scale scores in both Spring 2024 (baseline) and Spring 2025 (outcomes), grades 3-5.

iReady Diagnostic – Administered up to three times per year (Fall, Winter, Spring). The iReady panel (n=1,651) uses a flexible pre/post design: for each Rally Reader user, the baseline is their last iReady score before their Rally Reader start date, and the outcome is their most recent score after a minimum 180-day gap. Control students use the median Rally Reader start date (February 12, 2025) as their split point. This flexible design maximizes statistical power by leveraging each student's full exposure window rather than restricting to a fixed testing period.

Each assessment uses a different scale-scoring mechanism and scores are never compared directly.

Treatment Assignment

Students are classified into three treatment groups based on total Rally Reader usage hours relative to their grade's 75th percentile threshold:

Group	Criteria	n
Control	No Rally Reader Usage (< 2 min)	7,891
Exposure	Used Rally Reader (>=2 min)	1,222

Protocol	Used Rally Reader (P75)	379
----------	-------------------------	-----

The protocol threshold is computed per grade as the 75th percentile of total usage hours among students with any Rally Reader usage (>2 minutes). This within-grade approach accounts for the fact that typical usage levels vary by grade. P75 thresholds range from 0.6 hours (Grade 2) to 7.5 hours (Grade 6). A student is only classified as *treated* for a given assessment if their Rally Reader usage began *before* the assessment date.

In addition to the 3-arm model above, a binary "Any RR" model is estimated that pools all Rally Reader users (with at least 2 minutes of total usage) against all non-users, providing a simple intent-to-treat estimate.

Statistical Methods

Two primary analytical methods are used, each applied independently to both CAASPP and iReady panels:

1. OLS Difference-in-differences (DiD)
 1. Regresses score gain on treatment status, controlling for baseline score (centered within grade), grade level, gender, english learner status, economic disadvantage, and hispanic ethnicity. For iReady models, the number of days between baseline and outcome tests and an outcome window fixed effect are also included as controls to account for variation in the testing window and seasonal assessment differences. All standard errors are HC3-robust.
 2. Two OLS models are estimated: a 3-arm model (Exposure vs Protocol vs Control) for dose-response analysis, and a binary model (Any RR vs Control) for the pooled intent-to-treat estimate.
2. Propensity Score Matching + DiD
 1. A logistic regression model estimates the probability of Rally Reader usage based on the same covariates. Each treated student is matched 1:1 to a control student with the nearest propensity score (caliper = 0.2 SD). DiD is then computed on the matched pairs on matched assessments.
 2. PSM is run separately for two comparisons: Any RR user vs Control, and Protocol-only vs Control.

All p-values reported are from these regression models. Point estimates in tables reflect controlled regression estimates with full covariates.

Reading Type Classification

Rally Reader tracks whether students use oral (validated reading aloud) or silent reading modes. Students are classified by their share of oral reading time into three categories:

Classification	Criteria
Mostly Oral	>60% oral reading time
Mixed	40-60% oral reading time
Mostly Silent	<40% oral reading time

Demographic Controls

The following covariates are included in all regression models to control for observable differences between treatment and control groups:

1. Gender (50.0% male across iReady sample)
2. Hispanic ethnicity (47.4% of the sample)
3. English Learner status (6.7% of the sample)

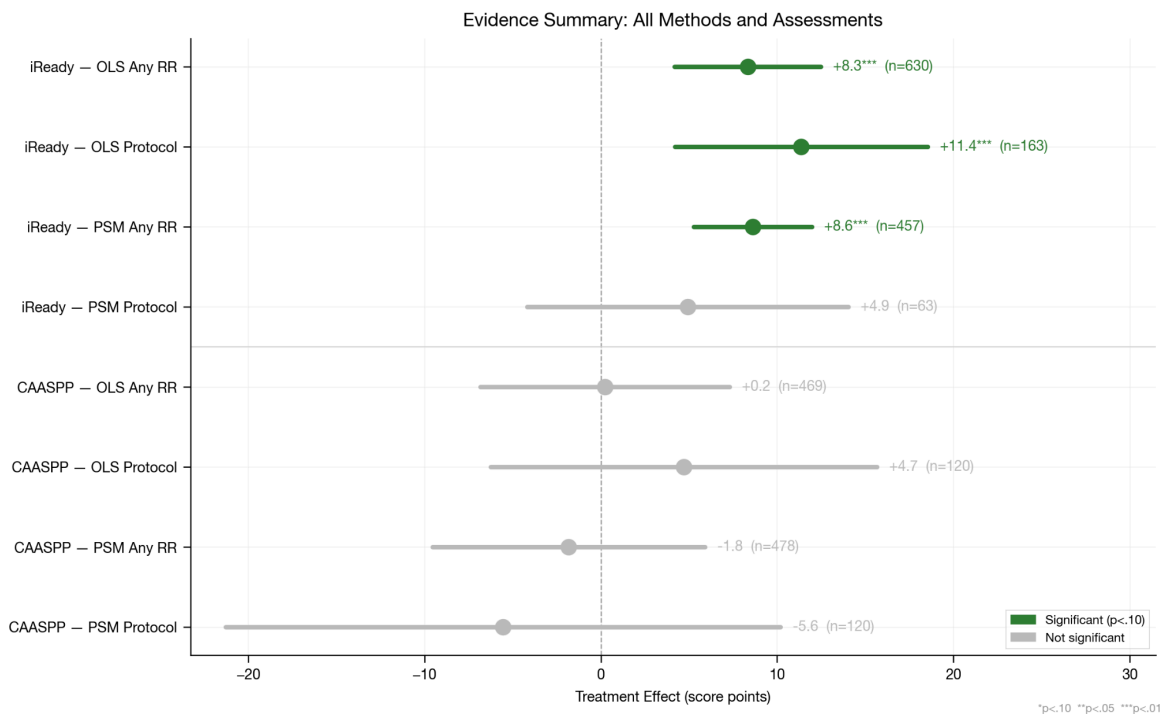
4. Economic disadvantage (38.5% of the sample)
5. Baseline Score (Centered within grade to control for starting achievement level)
6. Grade level (categorical fixed effect)
7. Days between tests (iReady models only — controls for variation in testing window)
8. Outcome window (iReady models only — fixed effect controlling for seasonal assessment differences)

The following variables were omitted due to a lack of sample size, availability, or introduction of significant expansion of degrees of freedom: Race, Special Education Status, Migrant status, School, Classroom.

Propensity score matching provides an additional check against OLS by explicitly balancing these covariates between treatment and control before estimating effects.

Results

All four iReady analyses — the more finely-measured assessment panel — indicate positive treatment effects for Rally Reader users, with three reaching statistical significance. Across the four CAASPP ELA analyses, results are directionally positive for protocol-adherent students, with Grade 5 reaching statistical significance. No evaluation produced a statistically significant negative result.



The strongest and most consistent evidence comes from the iReady diagnostics. The binary OLS estimate for all Rally Reader users is +8.3 points ($p < 0.001$), and the protocol-adherent OLS estimate is +11.4 points ($p = 0.002$). Propensity score matching on all users confirms the positive direction at +4.2 points ($p = 0.038$). CAASPP results are directionally positive for protocol students (+4.7 OLS) but carry wide confidence intervals due to smaller per-protocol sample sizes and are not statistically significant.

Overall Results: OLS DID and PSM (full demographic controls)

Assessment	Method	Arm	Estimate	95% CI	p	n(T)	n(C)
iReady	OLS	Any RR	+8.3***	[+4.2, +12.5]	0.0001	630	1021
iReady	OLS	Protocol	+11.4***	[+4.2, +18.5]	0.0019	163	1008
iReady	PSM	Any RR	+8.6***	[+5.3, +12.0]	0.0000	457	457
iReady	PSM	Protocol	+4.9	[-4.2, +14.1]	0.2882	63	63
CAASPP	OLS	Any RR	+0.2	[-6.8, +7.3]	0.9460	469	843
CAASPP	OLS	Protocol	+4.7	[-6.2, +15.6]	0.3997	120	834
CAASPP	PSM	Any RR	-1.8	[-9.6, +5.9]	0.6425	478	478
CAASPP	PSM	Protocol	-5.6	[-21.3, +10.2]	0.4890	120	120

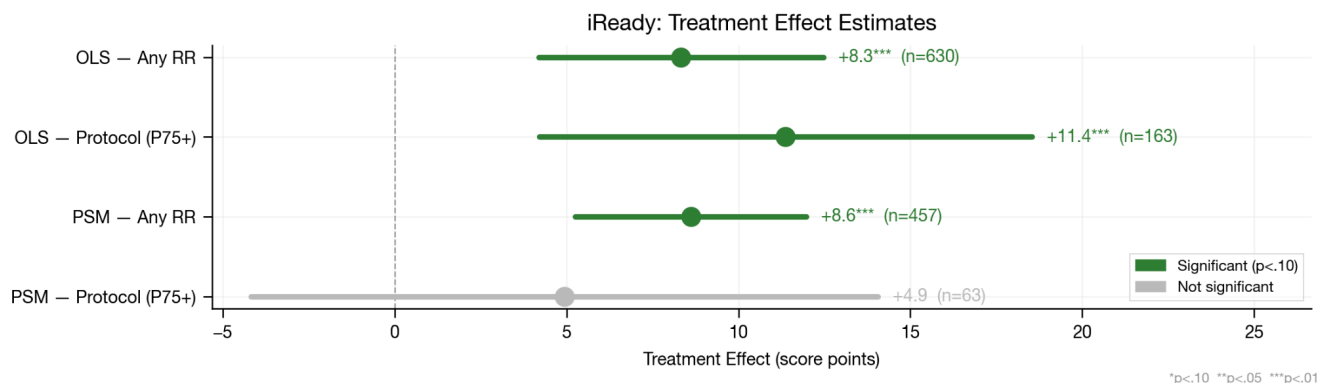
*p<.10 **p<.05 ***p<.01

iReady Diagnostic Results

Pooled Effects

Rally Reader users showed a significant improvement on iReady diagnostics relative to non-users. The *Any RR* binary OLS estimate of +8.3 points ($p < 0.001$, $n = 630$ users vs 1,021 control) represents the intent-to-treat effect across all usage levels.

Protocol-adherent students show a larger effect of +11.4 points ($p=0.002$, $n=163$), indicating a dose-response relationship. PSM confirms the positive direction for all users at +8.6 points ($p<0.001$, 457 matched pairs). PSM on protocol-only students yields +4.9 points ($p=0.288$, 63 pairs) – directionally positive but not significant given the smaller matched sample.



iReady: Primary Treatment Effect Estimates
Full demographic controls | % Extra Growth = DiD / control mean gain

Method	Treatment	n (Treat / Control)	DiD Effect	% Extra Growth	p-value
OLS	Any RR	630 / 1021	+8.3***	+43%	<0.001
OLS	Protocol (P75+)	163 / 1008	+11.4***	+59%	0.002
PSM	Any RR	457 / 457	+8.6***	+44%	<0.001
PSM	Protocol (P75+)	63 / 63	+4.9	+25%	0.288

*p<.10 **p<.05 ***p<.01

Per Grade Results

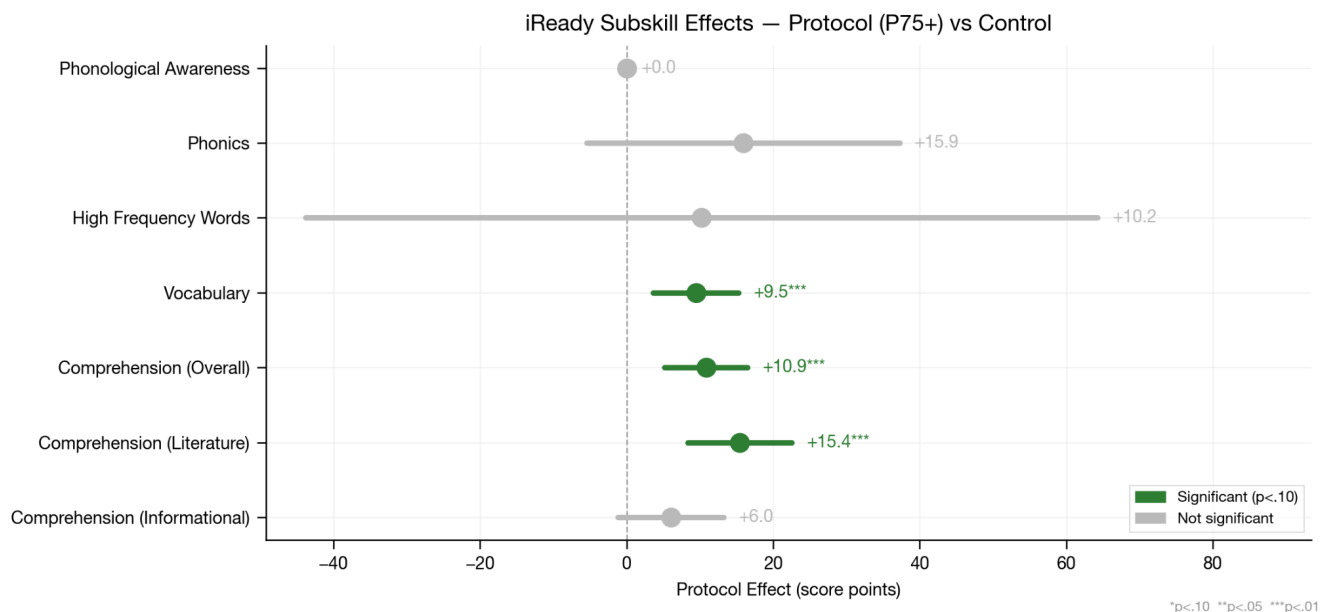
The treatment effect is strongest in grades 3 and 5. Protocol-adherent students in grade 5 show +20.0 points ($p=0.010$) and exposure students in grade 5 show +11.9 points ($p=0.001$). Grade 3 exposure students show +11.8 points ($p=0.043$), with protocol at +16.2 points ($p=0.087$). Grade 4 shows positive but non-significant effects.

iReady: Treatment Effect by Grade
Protocol (P75+) vs Control | % Extra Growth = DiD / control mean gain

Grade	n (Prot / Ctrl)	DiD Effect	% Extra Growth	p-value	% Oral
Grade 2	24 / 218	+11.0	+46%	0.229	85%
Grade 3	55 / 210	+16.2*	+132%	0.086	66%
Grade 4	42 / 114	+15.0	+79%	0.203	50%
Grade 5	31 / 193	+20.0***	+479%	0.010	47%

Subskills Analysis

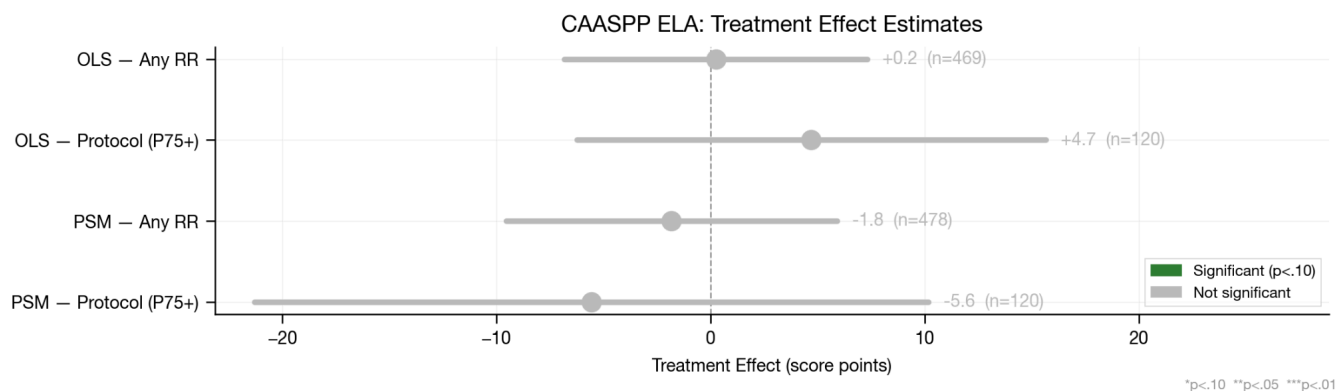
iReady subskill scores indicate that the treatment effect is significant in comprehension overall (+10.9 points), comprehension of literature (+15.4 points), vocabulary (+9.5 points), and comprehension of informational text (+8.8 points, $p=0.003$). We also see non-significant but directional gains in Phonics (+15.9 points) and High Frequency Words (+10.2) though this score is only available for lower grades. This is consistent with Rally Reader's content, which is primarily fiction-based material which caters to student interests.



CAASPP ELA Results

Pooled Effects

CAASPP results for protocol-adherent students are positive and directionally correlated with iReady but do not reach statistical significance at the pooled level. The OLS protocol estimate is +4.7 points ($p=0.40$, $n=120$), and PSM protocol yields -5.6 points ($p=0.49$, 120 matched pairs). The binary "Any RR" OLS estimate is near-zero (+0.2 points, $p=0.95$), indicating no pooled intent-to-treat effect on CAASPP. The CAASPP results carry wider confidence intervals due to the smaller per-protocol sample and the blunter nature of the annual assessment compared to iReady's adaptive and more frequent administration.



CAASPP ELA: Primary Treatment Effect Estimates
Full demographic controls | Regression-adjusted p-values

Method	Treatment	n (Treat / Control)	DiD Effect	p-value
OLS	Any RR	469 / 843	+0.2	0.946
OLS	Protocol (P75+)	120 / 834	+4.7	0.400
PSM	Any RR	478 / 478	-1.8	0.642
PSM	Protocol (P75+)	120 / 120	-5.6	0.489

Per Grade Results

The grade-level pattern for CAASPP shows a notable positive result for Grade 5 protocol students (+21.6 points, p=0.048). Other grade-level CAASPP results do not reach statistical significance.

CAASPP ELA: Treatment Effect and Oral Reading by Grade
Protocol (P75+) vs Control | Regression-adjusted p-values

Grade	n (Protocol / Control)	DiD Effect	p-value	% Oral Readers
Grade 03	57 / 193	-4.7	0.591	56%
Grade 04	34 / 300	+8.0	0.422	59%
Grade 05	29 / 341	+21.6**	0.048	34%

The divergence between CAASPP and iReady results may reflect assessment timing and frequency. CAASPP is administered only once per year in the Spring, while the iReady panel uses a flexible pre/post design based on each student's Rally Reader start date. Students who began Rally Reader mid-year may not have accumulated sufficient usage before the spring CAASPP assessment, or may have had a single poor test performance that would have been washed out in iReady's multiple testing windows.

Discussion

Interpretation

The results of this analysis provide reasonable converging evidence that Rally Reader usage has a positive effect on student reading achievement, particularly as measured by iReady diagnostics. The effect is robust across multiple analytical methods including OLS DiD, PSM DiD, and subskill decomposition. The indication of literary comprehension and vocabulary as the highest impact subskills lends credence to our hypothesis that Rally Reader is encouraging students to engage in reading through access to their preferred books and leading to improvements over their peers. There is also weak, but directional evidence that deeper engagement in oral reading, the primary reinforcement mechanism of Rally Reader, is leading to improved outcomes for students.

Dosage and Compliance

The protocol threshold in this analysis is defined as the within-grade 75th percentile (P75) of total usage hours, rather than a fixed hour count. This data-driven approach identifies students whose engagement is meaningfully above their grade peers and accounts for natural differences in usage patterns across grades. Protocol thresholds range from 0.9 hours in Grade 2 to 9.7 hours in Grade 6.

Additional analysis is warranted to examine the relationship between total reading time and academic improvement. We did not observe a linear correlation between increased dosage and additional performance gains. This pattern is consistent with

a ceiling effect rather than diminishing returns from the product itself: students who read the most tend to be higher-performing to begin with, and high-achieving students have less room to gain on assessments calibrated to grade-level performance.

Dosage compliance, which is defined as an aggregate usage over the school year, was an opinionated decision made for this analysis and is subject to scrutiny. This was chosen over alternatives of average time per week, active weeks, active days, and average time per session after assessing the sample sizes and baseline demographics of students in the exposure group. We found that a simple aggregate number of reading hours was a straightforward measure of dosage compliance and allowed a reasonable number of students in the sample rather than a stricter protocol of 45 minutes each week during the school year. However improving the acceptance protocol selection and dosage compliance in a future study will allow much more accurate measurement.

Limitations

There are several key limitations of this study that must be addressed.

1. Single District
 - 1.1. All data comes from the school district. Results may not generalize to districts with different demographics, adoption patterns, or instructional contexts.
2. Small per-protocol samples
 - 2.1. Protocol sample sizes vary by assessment panel (154 in iReady, 120 in CAASPP). Per-grade protocol samples are smaller still, which limits the precision of per-grade estimates and contributes to wide confidence intervals.
3. Non-random assignment
 - 3.1. Adoption was teacher and student driven and largely voluntary. While this study attempts to control for differences in exposure vs control, unobserved confounders such as teacher quality, school, or classroom level effects cannot be ruled out.
4. Flexible panel design
 - 4.1. The iReady panel uses each student's Rally Reader start date to define the pre/post split, which maximizes statistical power but introduces variation in the testing window. Because the flexible design produces outcomes spanning multiple seasonal iReady windows (Fall, Winter, Spring), an outcome window fixed effect is included to control for seasonal re-norming differences. Days between tests are also controlled for in the regression. A same-window robustness check restricted to Fall 2025 outcomes confirms the positive result (Any RR +7.5, $p=0.005$; PSM +5.7, $p=0.022$). A Spring-to-Spring sensitivity analysis (fixed window) shows directionally consistent but smaller, non-significant effects.

Next Steps

One of the key outcomes of this post-hoc analysis is to identify what we may require in a future randomized controlled trial to more clearly identify the effects of Rally Reader on student reading achievement. There are a few concerns, namely how many students we need in the treatment group to get a significant number of students to reach dosage compliance and how many students we need in the treatment group to identify a potentially lower treatment effect.

With a conservative treatment effect of +3-4pts we would need a total of 4,036 students to measure a statistically significant effect in CAASPP scores. Reasonable RCTs would likely need somewhere between 1500 and 4000 students appropriately randomized and controlled. We would also need to make protocol adherence monitored and enforced as part of the study rather than chosen after the fact.

Appendix

Appendix A Regression Coefficients

Appendix A: iReady Full Regression | R²=0.164, N=1651

Variable	Coef.	SE	p
Intercept	+21.92***	5.42	0.0001
Grade 2	-8.50***	2.64	0.0013
Grade 3	-19.50***	2.54	0.0000
Grade 4	-17.85***	2.78	0.0000
Grade 5	-26.92***	2.61	0.0000
Grade 6	-31.16***	6.09	0.0000
C(outcome_window)[T.Spring (March 2 - End of Year)	-2.11	3.94	0.5922
C(outcome_window)[T.Winter (November 16 - March 1)	-7.12***	2.64	0.0070
Exposure	+7.46***	2.14	0.0005
Protocol (P75+)	+11.37***	3.66	0.0019
Baseline Score (ctrd)	-0.16***	0.02	0.0000
Econ. Disadvantaged	-3.84**	1.58	0.0148
English Learner	-5.91	4.22	0.1613
Hispanic	-3.07**	1.51	0.0418
Male	-1.57	1.45	0.2809
Days Between Tests	+0.06***	0.02	0.0053

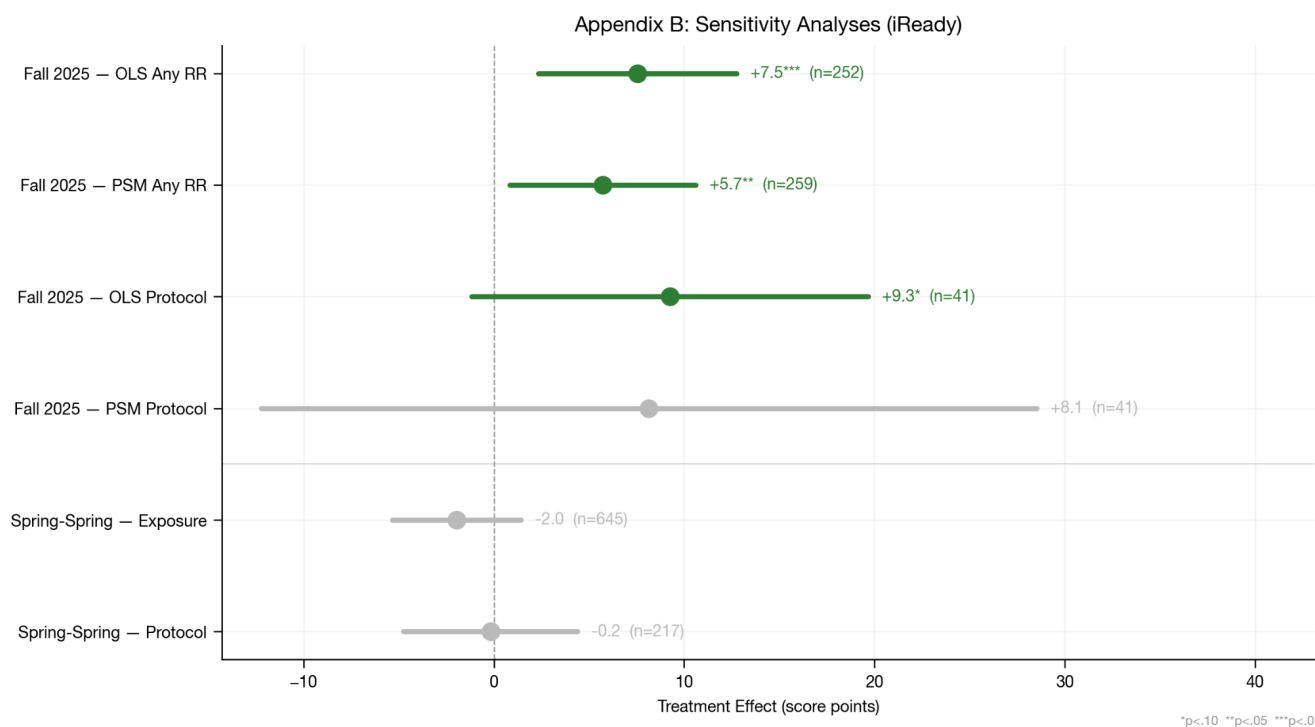
*p<.10 **p<.05 ***p<.01

Appendix A: CAASPP Full Regression | R²=0.137, N=1312

Variable	Coef.	SE	p
Intercept	+65.01***	4.10	0.0000
Grade 04	-11.69***	4.10	0.0044
Grade 05	-35.24***	4.08	0.0000
Protocol (P75+)	+5.01	5.83	0.3899
Baseline Score (ctrd)	-0.21***	0.02	0.0000
Econ. Disadvantaged	-7.92**	3.57	0.0266
English Learner	-22.89***	8.23	0.0055
Hispanic	-5.16	3.40	0.1296
Male	-4.01	3.31	0.2265

*p<.10 **p<.05 ***p<.01

Appendix B Sensitivity



Because a fixed Spring-to-Spring window compresses the effective treatment period — it includes students who began Rally Reader after their baseline Spring test but before the outcome Spring test — attenuated effects are expected in this sensitivity analysis relative to the flexible panel. A Spring-to-Spring robustness panel restricts the iReady analysis to students with Spring 2024 and Spring 2025 scores (n=1,562). Under this fixed-window design, the protocol estimate is -0.2 points and the exposure estimate is -2.0 points (p=0.43). Neither reaches significance.

Appendix C PSM Covariate Balance

Appendix C: PSM Balance (iReady Protocol)
162 matched pairs | Green = SMD < 0.15

Covariate	Treat Mean	Control Mean	Pre SMD	Post SMD
Baseline Score (std)	0.475	-0.052	+0.580	-0.012
Econ. Disadvantaged	0.233	0.403	-0.369	-0.043
English Learner	0.037	0.071	-0.150	-0.061
Hispanic	0.362	0.486	-0.252	+0.091
Male	0.436	0.507	-0.144	+0.075

*p<.10 **p<.05 ***p<.01