# 95 PHONICS CORE PROGRAM ${ }^{\text {TM }}$ 

 FALL 2021, GRADES K-2 EFFICACY STUDY
# 95 Phonics Core Program ${ }^{\text {TM }}$ Classroom Kit Grades K-3 

Efficacy Study with Acadience ${ }^{\circledR}$ Reading K-6: Level 1 ESSA Level of Evidence Conducted by Rachel Schechter, Ph.D. and LXD Research at Charles River Media Group

## Executive Summary

There is a growing concern that core reading curricula for the elementary years have not improved reading scores in the US (The Condition of Education 2020). Reporters such as Emily Hanford (APM Reports) have shone a light specifically on the need for explicit, systematic, and sequential phonics instruction for every child. In response to this identified need, 95 Percent Group, LLC created a new phonics core curriculum that can replace the phonics instructional lessons provided with other core reading curricula (typically the first 20 minutes of the reading block).

The 95 Phonics Core Program (95PCP) is a whole-class, Tier I program designed for students in grades K-3 to address and prevent reading gaps using explicit, structured phonics instruction for 20 minutes per day. Instruction is based on a scope and sequence with 25 lessons for kindergarten and 30 lessons for each of Grades 1-3. The 95 Percent Group hired LXD Research, a third-party independent evaluator, to conduct a research study on the impact of the 95 PCP .

## Research Description

The research study has one primary goal: to examine the impact of the 95PCP on student literacy achievement in Grades K-3. Additional goals of the study are to examine program implementation information and feedback from educators about the program's quality and ease-of-use, as well as their perceptions about the impact of the 95PCP on teaching and learning. To meet these goals, LXD Research designed a mixed-methods study with random selection of participants into conditions at the school level (ESSA Level 1 - Strong). This paper focuses on grades K-2 for the first semester of implementation only.

The recruited research partner is in a majority-White school district in Missouri. There are over 6,000 students in grades K-5 across 17 elementary schools in the district. The district serves a population in which $23 \%$ of students traditionally ${ }^{1}$ qualified for free lunch and between $5-12 \%$ of students per grade are English Language Learners (described later as "EL status"). LXD Research is engaging with the district to conduct a set of qualitative and quantitative data collection activities, of which two main activities were completed in Fall and Winter 2021: Acadience Reading K-6 assessment administration

[^0](at the Beginning of Year, BOY, and the Middle of Year, MOY) and a teacher survey with control schools (the teacher survey for treatment schools will be administered in the Spring of 2022).

Research Questions and Methods

## Sample > Design > Analysis

- Schools with similar characteristics were paired
- Schools within each pair were randomly assigned to treatment (adds 95PCP) or control (does not add 95PCP)
- Schools with 95PCP replace ReadyGEN phonics time with 95PCP
- Control schools use ReadyGEN and other "business-as-usual" materials
- Students in all schools assessed with Acadience Reading
- Teacher survey
- Acadience Reading scores analyzed by grade to understand differences between groups
- Advanced statistical modeling accounts for multiple characteristics


## Research Questions

- How does the 95PCP affect K-2 student achievement on formative assessments (of phonics, specifically) in schools that implement the program compared to schools that do not implement the program by MOY?
- How does the impact of the 95PCP vary by grade and student subgroup (gender, English Language Learner status, students in special education)?
- How does the impact of the 95PCP vary by a student's BOY benchmark status (i.e., do students Below Benchmark at BOY achieve similar growth as students who were At Benchmark at BOY)?
- What, if any, impact does teacher training in the Science of Reading have on student scores?

An advanced statistical modeling analysis was conducted to understand what the data show in support of answering these questions. These techniques help account for known differences that could impact student outcomes.


Fall 2021 Results Highlights (see full paper for all results for all research questions)


## Student BOY Benchmark Status

After only 14 weeks of use, the data show that a positive impact of the 95 Phonics Core Program was measurable for all three grades. Positive impacts were found for the following Acadience Reading measures:

- Grade K: Composite, FSF (First Sound Fluency), and LNF (Letter Naming Fluency)
- First Grade: Nonsense Word Fluency CLS (Correct Letter Sounds) and WWR (Whole Words Read)
- Second Grade: Composite and ORF (Oral Reading Fluency)

Kindergarten: Treatment group showed more growth from BOY to MOY than the Control group in Composite Scores


Kindergarten: Treatment group showed more growth from BOY to MOY than the Control group in FSF (First Sound Fluency) and LNF (Letter Naming Fluency) Scores


First Grade: Treatment group showed more growth from BOY to MOY than the Control group in Nonsense Word Fluency CLS (Correct Letter Sounds) and WWR (Whole Words Read) Scores


## Second Grade: Treatment group showed more growth from BOY to MOY than the Control group in Composite and ORF (Oral Reading Fluency) Scores




## Conclusion \& Next Steps

While the 95PCP is intended to be used all year, the positive impacts during the Fall semester (14 weeks) are encouraging and support the ESSA Level 1 evidence for the 95PCP's effectiveness. The analysis of the MOY Acadience data collection begins to answer the central questions of this research study. In particular, the following data analysis findings provide reasonable and positive answers to the research questions in advance of our EOY data collection:

- How does the 95PCP affect K-2 student achievement on formative assessments (of phonics, specifically) in schools that implement the program compared to schools that do not implement the program by MOY?
- The data show a positive, significant impact on student Acadience scores for all grades (K-2) on multiple subtests, as well as on the Composite score for grades K and 2.
- How does the impact of the 95PCP vary by grade and student subgroup (gender, English Language Learners, students in special education)?
- Students performed similarly regardless of gender, EL status, or SPED status.
- How does the impact of the 95PCP vary by a student's BOY benchmark status (i.e., do students Below Benchmark at BOY achieve similar growth as students who were At Benchmark at BOY)?
- Students performed similarly regardless of BOY benchmark status. In other words, students who were Below Benchmark or Well Below Benchmark made the same amount of progress as students who started the year At or Above Benchmark.
- What, if any, impact does teacher training in the Science of Reading (SOR) have on student scores?
- The impact of teacher SOR training is not straightforward. For kindergarten, the training seemed to support student growth; in first grade the data show students of teachers in the training made fewer gains; and in second grade there was no apparent impact. Reasonable hypotheses about why first graders made fewer gains may include the impact of time scarcity for teachers simultaneously participating in the SOR course and implementing a new phonics program. There was no detectable impact of teacher training on the learning of students in the control schools.

As the study continues during the Winter of 2022, Administrator interviews will allow for a better understanding of how phonics instruction and intervention has progressed over the course of the first half of the year. These more in-depth conversations will provide additional context from both treatment and control schools on what instruction looks like across grades and within each building. The Spring data collection includes focus groups with educators and the end-of-year (EOY) Acadience data collection.

## Table of Contents

InTRODUCTION ..... 2
Evaluation Questions ..... 2
Methods ..... 2
Design ..... 3
Treatment Group: Key Program Features ..... 3
Control Group: Phonics Instruction ..... 5
Control Group: Survey Summary ..... 6
Assessment: Acadience Reading K-6 ..... 6
SAmple ..... 6
Analytic Approach ..... 8
Results ..... 9
Kindergarten ..... 9
First Grade ..... 10
SEcond Grade ..... 11
Conclusion \& Next Steps ..... 11
Appendices ..... 13
Appendix 1: Kindergarten Results ..... 13
Appendix 2: First Grade Results ..... 21
Appendix 3: Second Grade Results ..... 29

## Introduction

There is a growing concern that core reading curricula for the elementary years have not improved reading scores in the US (The Condition of Education 2020). Reporters such as Emily Hanford (APM Reports) have shone a light specifically on the need for explicit, systematic, and sequential phonics instruction for every child. In response to this identified need, 95 Percent Group, LLC created a new phonics core curriculum that can replace the phonics instructional lessons provided with other core reading curricula (typically the first 20 minutes of the reading block).

The 95 Phonics Core Program (95PCP) is a whole-class, Tier I program designed for students in grades K-3 to address and prevent reading gaps using explicit, structured phonics instruction for 20 minutes per day. Instruction is based on a scope and sequence with 25 lessons for Kindergarten and 30 lessons for each of Grades 1-3. For example, the First Grade Scope and Sequence includes 30 lessons disaggregated into seven topics (introduction, short vowel CVC, consonant blends, consonant digraphs, long vowel silent-e, phonograms, and introduction to second-grade skills). Each lesson focuses on specific phonics skills, provides examples of high-frequency words, and contains information about other skills addressed within the topic. The 95 Percent Group offers a kit for each grade, including a teacher's edition, student workbooks, manipulatives, and a digital presentation. The 95PCP may be offered in-person or virtually. The 95PCP also aligns with assessments and interventions (such as Phonics Lesson Library) offered by 95 Percent Group to ensure consistency.


95 Percent Group partnered with LXD Research to conduct a third-party evaluation of the 95PCP as it was implemented during the 2021-2022 school year in a Missouri school district. All the elementary schools use ReadyGEN as a core reading curriculum and half were randomly selected to use the 95PCP for phonics instruction instead of the ReadyGEN word study materials. Random assignment to conditions ensures the highest level of scientific rigor (ESSA Evidence Level 1).

## Evaluation Questions

The evaluation aims to answer the following questions:

1. How does the 95PCP affect student achievement on formative assessments (of phonics, specifically) in schools that implement the program compared to schools that do not implement the program?
2. How does the impact of the 95PCP vary by school, grade, and student subgroup (gender, English Language Learners [EL] status, students in special education [SPED] status)?
3. What is the nature and extent of the 95PCP implementation?
a. How is the 95PCP typically implemented?
b. To what extent is the 95PCP implemented with fidelity, and does the program adhere to the Theory of Action?
c. How do contextual factors affect 95PCP implementation, such as the content and quality of professional development and the characteristics of districts and schools, such as the level of administrator support?
4. What is the nature and extent of literacy program implementation in comparison schools?
5. What are teacher and administrator perceptions about the quality and impact of the 95 PCP ?
d. What are teacher and administrator initial reactions to the 95 PCP and its associated materials, content, pacing, and professional development?
e. What suggestions do they have for improvement?
6. What is the association between variations in 95 PCP implementation and student outcomes?

## Methods

This study uses a mixed-method design that includes quantitative and qualitative data collection. The Fall 2021 and Winter 2022 research activities included the beginning-of-year (BOY) reading assessment, a middle-of-year (MOY) assessment, and a survey of teachers from control schools that focused on phonics instructional strategies and background.

## Design

This study uses a mixed-methods approach, including a randomized experimental design complemented by classroom observations, teacher surveys, and administrator interviews. This combination of methods allows researchers to understand how the materials are being used in the classroom, collect teacher feedback on the quality and perceived impact of the program, and evaluate student academic achievement.

School districts were recruited in Spring 2021. In exchange for their participation, district leaders received all 95PCP materials and training at no cost and discounts for any 95PCP materials purchased in the 2022-2023 school year. The control schools used the regular materials that they have used in previous years. Prior to the 2021-2022 school year, the district leaders allowed for the randomization of schools to a treatment ( 95 PCP ) or control condition. Schools were organized into pairs using school size and ELA scores from Fall 2020 and Spring 2021, and then a coin toss determined which school in each pair would receive the 95PCP. Students were pretested within the first four weeks of school using Acadience Reading (BOY: September 1-17²) and were tested again between December 13-17, 2021 (middle of the year, or MOY, after at least 12 weeks of instruction). The final testing period will be at the end of the year (EOY) in Spring 2022.

## Treatment Group: Program Key Features

The 95PCP features instructional practices that differ from the typical reading instruction provided by core curricula. A phonemic awareness and phonics continuum of skills is followed using structured literacy characteristics, described in Table 1.

[^1]Table 1.95 Percent Group's Literacy Cbaracteristics in 95PCP Lessons

| Characteristic | Evident in Lesson Framework |
| :--- | :--- |
| 1. Explicit | IDo directly states and defines focus skill and student expectations. |
| 2. Systematic | Intentional language and steps include consistent hand gestures and <br> verbal cues; there is a gradual transfer of responsibility from teacher to <br> student. |
| 3. Sequential | Structure moves from simple to complex in key ways including <br> lesson order, word choice, materials used, and teacher talk. |
| 4. Adequate Modeling | This most prominent feature provides precise language at each level <br> of modeling. |
| 5. Corrective Feedback | Teacher response is reactive to individual student errors. |
| 6. Differentiated <br> Instruction | We Do and You Do sections provide two levels that enable teachers <br> to differentiate instruction to meet students' needs. |
| 7. Scaffolded <br> Instruction | Steps of the I Do, We Do, and You Do allow the teacher to gradually <br> transfer responsibility for learning to the students. |
| 8. Continual <br> Assessment | This occurs through informal observation and monitoring during <br> instruction; the focus skill correlates to the PSI. |

The 95 Percent Group's version of the gradual release model (Table 2) allows all students to practice every skill using multisensory materials, including a phonics mat and chips. While a paper version of the Phonics Chip Kit is included in the 95PCP, a plastic version is available and sold separately.

## Table 2. Gradual Release Model in 95PCP

| Modeling Steps | Chip Movement | Speaking |
| :--- | :--- | :--- |
| I Do | teacher | teacher |
| We Do |  |  |
| Level 1: Accuracy | teacher | teacher and students |
| Level 2: Fluency | teacher | students |
| You Do | students | students |

The 95PCP phonological awareness and phonics continua are shown in Figures 1 and 2, respectively. There is a clear progression from simpler to more complex skills, following the research-based developmental progression for learning to read. The International Dyslexia Association, for example, describes structured literacy as "Systematic means that organization of material follows the logical order of language. The sequence begins with the easiest and most basic concepts and elements and progresses methodically to the more difficult."

Figure 1. Phonological Awareness Continuum of 95 Percent Group


Figure 2. Phonics Continuum of Skills of 95 Percent Group


## Control Group: Phonics Instruction

The district uses ReadyGEN for their core reading program, which is published by Savvas Learning Company (formerly Pearson). This curriculum has one published study that meets the Level 3 (Promising) ESSA criteria for first grade using the Terranova 3 assessment. The program is described as using the Gradual Release of Responsibility Model, a generative approach to vocabulary instruction, and many language-focused, text-based strategies for teaching reading and writing. The curriculum also includes assessments and online games.

## Control Group: Survey Summary

A survey conducted in the Fall of 2021 collected information from the teachers in the control schools about their approach to teaching phonics across all tiers. A total of 83 teachers who teach K-3, the focus grades for the 95 PCP , responded; the respondents included at least one representative from each school. Nearly all teachers use ReadyGEN to teach phonics, with Phonics First being the other consistently mentioned program. More than $25 \%$ of teachers indicated that ReadyGEN didn't have phonics instruction, or they didn't know if it did. The most-used supplemental phonics program was Reading A-Z, with Heggerty Phonemic Awareness and Phonics First following behind. A small group of teachers mentioned using 95 Percent Group's Multisyllable Routine Cards (Second Grade only), Heggerty Bridge the Gap, Raz-Kids, and Leveled Literacy Intervention materials. The amount of time allocated for phonics instruction varied widely within schools and grades, ranging from no time to more than 25 minutes per day. A survey of the teachers in the treatment schools will be conducted in Spring 2022.

## Assessment: Acadience Reading K-6

Acadience Reading is an assessment that helps teachers identify children at risk for reading difficulties and determine the skills to target for instructional support. Acadience assessments are standardized and assess core early literacy skills (Table 3). Because the subtests and their weighting change for each assessment period (see Acadience User Manual), Composite scores are used to compare reading ability in this report.

## Table 3. Acadience Reading Subtests and Skill Coverage

| Subtest | Indicators of These Basic Early Literacy Skills |
| :--- | :--- |
| First Sound Fluency (FSF) \& Phoneme <br> Segmentation Fluency (PSF) | Phonemic Awareness |
| Letter Naming Fluency (LNF) | Indicator of risk |
| Nonsense Word Fluency (NWF) | Alphabetic Principle and Basic Phonics (Correct Letter Sounds and <br> Whole Words Read) |
| Oral Reading Fluency (ORF) \& Retell <br> Fluency (RTF) | Advanced Phonics and Word Attack Skills, Accurate and Fluent <br> Reading of Text (ORF Words Correct Per Minute and Errors); <br> Reading Comprehension (RTF Total and Quality of Response) |
| Maze | Reading Comprehension |

## Sample

The 95PCP is being implemented in a majority-White school district in Missouri. The district serves a population in which $23 \%$ of students traditionally ${ }^{3}$ qualified for free lunch. A total of 3,327 K-2 students in 178 classrooms across 14 elementary schools were included in this analysis, further described in Table 4.

Table 4. Number of Students, Classes, and Schools by Grade and Condition

| Grade Level | School Group | \# of Classes | BOY | MOY | Matched Sample |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  |  |  | \# of Students | \# of Students |  |
|  | Treatment | 32 | 612 | 599 | 598 |
|  | Control | 27 | 517 | 497 | 497 |
|  | Total | 59 | 1129 | 1096 | 1095 |
| $\mathbf{2}$ | Treatment | 30 | 578 | 567 | 565 |
|  | Control | 29 | 532 | 519 | 519 |
|  | Total | 59 | 1110 | 1086 | 1084 |
|  | Treatment | 33 | 631 | 624 | 623 |
|  | Control | 27 | 536 | 527 | 525 |
|  | Total | 60 | 1167 | 1151 | 1148 |

Of these students, 1,786 were in the treatment group and 1,541 were in the control group. Students in the treatment and control groups were similar with respect to gender and SPED status. However, students in the control group were more likely to be ELLs ( $\mathrm{B}=-.03$, $\mathrm{p}<.001$; see Table $5^{*}$ ).

Table 5. Demographic descriptions for treatment and control group

| Group | Male | SPED | ELL* $^{*}$ |
| :--- | :--- | :--- | :--- |
| Control | $49 \%$ | $14 \%$ | $10 \%$ |
| Treatment | $50 \%$ | $13 \%$ | $7 \%$ |

## Acadience Reading Beginning-of-Year Scores

The random assignment of schools successfully created similar treatment and control groups in each grade. The significance level for each pair was greater than 0.05 , and effect sizes (Hedges'g) were 0.11 or lower for all grades (Table 6).

[^2]Table 6. Acadience Composite Score Results for Beginning of Year (all students with BOY)

| Grade | Condition | Number of students | Treatment avg | SD | Significance | Effect Size Cohen's d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| K | Treatment | 612 | 31.69 | 23.98 | $\mathrm{p}=.06$ | . 12 |
|  | Control | 517 | 29.04 | 23.78 |  |  |
| 1 st grade | Treatment | 578 | 99.24 | 42.21 | $\mathrm{p}=.81$ | . 01 |
|  | Control | 532 | 99.85 | 40.21 |  |  |
| 2nd grade | Treatment | 631 | 162.66 | 87.59 | $\mathrm{p}=.04$ | . 06 |
|  | Control | 534 | 152.51 | 85.29 |  |  |

## Analytic Approach

This report focuses on exploring the following research questions:

- How does the 95PCP affect K-2 student achievement on formative assessments (of phonics, specifically) in schools that implement the program compared to schools that do not implement the program by MOY?
- How does the impact of the 95PCP vary by school, grade, and student subgroup (gender, EL, and SPED status)?
- How does the impact of 95PCP vary by a student's BOY benchmark status (i.e., do students Below Benchmark at BOY achieve similar growth as students who were At Benchmark at BOY)?
- What, if any, impact does teacher training in the Science of Reading have on student scores?

To answer these questions, three-level hierarchical linear regression models (HLMs) with time (level 1) nested within students (level 2) nested with classrooms (level 3) were employed to examine growth in Acadience Reading Composite and subscale scores. All models contained a series of covariates including gender ("Gender"; $1=$ male, $0=$ female), ELL status ("ELL"; $1=E L L, 0=$ non-ELL), SPED status ("SPED"; $1=$ SPED, $0=$ non-SPED), an indicator of time ("Time"; $1=\mathrm{BOY}, 2=\mathrm{MOY}$ ), an indicator of whether the student was in the treatment or control group ("group"; $1=$ Control, $2=$ Treatment), and an interaction between time and group calculated as the product of time*group ("Tigr").

We explored main effects of treatment versus control group by considering the significance of the interaction between time and group ("Tigr"). A significant interaction term would suggest that the slope (i.e., growth) in Composite scores is different for the treatment versus control groups. We also looked at growth in Composite scores separately based on students' BOY benchmark status. Finally, we considered whether formal teacher training ("sort"; $0=$ No Training, $1=1+$ Years Training) in the Science of Reading (not a 95 Percent Group program) moderated growth in Composite scores. All analyses were conducted separately by grade using the statistical software package R 3.6.2.

## Results

## Kindergarten

Within the Kindergarten grade sample, we examined growth in Composite scores as well as growth in First Sound Fluency (FSF) and Letter Naming Fluency (LNF) scores. Because the scores were highly negatively skewed counts, we elected to use a Poisson distribution to examine changes in scores over time. There was a significant effect of treatment on Composite scores (IRR=.98, p=.044), FSF scores ( $\operatorname{IRR}=.96, \mathrm{p}=.032$ ), and LNF scores ( $\mathrm{IRR}=-.97, \mathrm{p}=.039$ ) Figures 3a-c show these effects graphically. There were no significant findings regarding benchmark status. That is, students tended to demonstrate similar growth in scores regardless of their benchmark status at BOY. While there was no significant effect of teacher training within the control group, there was a significant effect of teacher training within the treatment group ( $\operatorname{IRR}=.94, \mathrm{p}=.005$ ), suggesting that treatment students whose teachers had training in the Science of Reading tended to demonstrate more growth in Composite scores than treatment students whose teachers had no training (Figure 3d). Complete output for each model can be found in Appendix 1.

Figure 3a. Kindergarten students in the treatment group demonstrated significantly more growth in Composite scores than students in the control group


Figure 36. Kindergarten students in the treatment group demonstrated significantly more growth in FSF scores than students in the control group

K FSF Scores BOY-MOY by Group


Figure 3c. Kindergarten students in the treatment group demonstrated significantly more growth in LNF scores than students in the control group

K LNF Scores BOY-MOY by Group


Figure 3d. Kindergarten treatment group students whose teachers had some training in the Science of Reading demonstrated significantly more growth in Composite scores than students whose teachers had no training

## K Composite Scores, PCP Group, by Teacher SOR Training



## First Grade

Within the First-Grade sample, we examined growth in Composite scores as well as growth in Nonsense Word Fluency Correct Letter Sound (CLS) and Whole Words Read (WWR) scores. Students in the treatment group demonstrated significantly more growth in CLS ( $\mathrm{B}=2.50, \mathrm{p}=.047$ ) and WWR ( $\mathrm{B}=1.38, \mathrm{p}=.019$ ) scores than students in the control group, as shown in Figures 4 a and 4 b , respectively. There were no statistically meaningful differences between the treatment and control groups regarding Composite scores. There were also no significant findings regarding benchmark status. That is, students tended to demonstrate similar growth in scores regardless of their benchmark status at BOY. While there was no significant effect of teacher training within the control group, there was a significant effect of teacher training within the treatment group ( $B=-14.06, p=.025$ ), suggesting that treatment group students of teachers who had training in the Science of Reading tended to demonstrate less growth in Composite scores than those whose teachers had no training (Figure 4c). Complete output for each model can be found in Appendix 2.

Figure 4a. First Grade students in the treatment group demonstrated significantly more growth in CLS scores than students in the control group


Figure 4c. Treatment group students whose teachers had some training in the Science of Reading demonstrated significantly less growth in Composite scores than students whose teachers had no training

Grade I Composite Scores, PCP Group, by Teacher SOR Training


## Second Grade

Within the Second-Grade sample, we examined growth in Composite scores as well as improvement in Oral Reading Fluency Words Correct Per Minute (ORF) scores, Oral Reading Fluency Error (ERR) scores, Retell Total (RETELL) scores, and Retell Quality (RETELLQR) scores. Students in the treatment group demonstrated significantly more growth in Composite scores $(\mathrm{B}=7.17, \mathrm{p}=.024$, $\mathrm{f}^{2}=.01$ ) and OFR scores $\left(\mathrm{B}=3.93, \mathrm{p}<.001, \mathrm{f}^{2}=.01\right)$ than students in the control group (Figures $5 \mathrm{a}-\mathrm{b}$ ). There were no statistically meaningful differences between the treatment and control groups regarding ERR, RETELL, or RETELLQR scores. There were also no significant findings regarding benchmark scores or teacher training. That is, students tended to demonstrate similar growth in scores regardless of their benchmark status or whether their teacher had SOR training. Complete output for each model can be found in Appendix 3.

Figure 5a. Second Grade students in the treatment group demonstrated significantly more growth in Composite scores than students in the control group

## Second Grade Composite Scores BOY-MOY by Group



Figure 5b. Second Grade students in the treatment group demonstrated significantly more growth in ORF scores than students in the control group

## Second Grade ORF Scores BOY-MOY by Group



- How does the 95PCP affect K-2 student achievement on formative assessments (of phonics, specifically) in schools that implement the program compared to schools that do not implement the program by MOY?
- The data show a positive, significant impact on student achievement on multiple Acadience subtests for all grades (K-2), as well as on the Composite scores for grades K and 2.
- How does the impact of the 95PCP vary by grade and student subgroup (gender, EL, or SPED status)?
- Students performed similarly, regardless of Gender, EL or SPED status.
- How does the impact of the 95PCP vary by a student's BOY benchmark status (i.e., do students Below Benchmark at BOY achieve similar growth as students who were At Benchmark at BOY)?
- Students performed similarly, regardless of BOY benchmark status. In other words, students who were Below Benchmark or Well Below Benchmark made the same amount of progress as students who started the year At or Above Benchmark.
- What, if any, impact does teacher training in the Science of Reading have on student scores?
- The impact of teacher SOR training is not straightforward. For kindergarten, the training seemed to support student growth; in first grade the data show students of teachers in the training made fewer gains; and in second grade there was no apparent impact. Reasonable hypotheses about why first graders made fewer gains may include the impact of time scarcity for teachers simultaneously participating in the SOR course and implementing a new phonics program. There was no detectable impact of teacher training on the learning of students in the control schools.

As the study continues during the Winter of 2022, Administrator interviews will allow for a better understanding of how phonics instruction and intervention has progressed over the course of the first half of the year. These more in-depth conversations will provide additional context from both treatment and control schools on what instruction looks like across grades and within each building. The Spring data collection includes focus groups with educators and the EOY Acadience data collection.

## Appendices

## Appendix 1: Kindergarten Results

- Composite score: (IRR=0.98, $\mathrm{p}=0.044$ ) - significant differences between treatment and control groups
- First Sound Fluency (FSF) score: (IRR=0.96, p=0.032) - significant differences between treatment and control groups
- Letter Naming Fluency (LNF) score: (IRR=-0.97, $\mathrm{p}=0.039$ ) - significant differences between treatment and control groups

For Below or Well Below Benchmark students:

- Composite score: (IRR=-2.14, $\mathrm{p}=0.343$ ) - no significant differences between treatment and control groups

For Below Benchmark students:

- Composite score: (IRR=-1.00, $\mathrm{p}=0.932$ ) - no significant differences between treatment and control groups


## Exploring teacher training:

- EXPERIMENTAL GROUP: Composite score: $(\operatorname{IRR}=0.94, \mathrm{p}=0.005)$ - significant effect of teacher training
- CONTROL GROUP: Composite score: (IRR=1.03, $\mathrm{p}=0.184$ ) - no effect of teacher training

|  | ark 6 comp K |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Incidence Rate Ratios | $C I$ | $p$ |
| (Intercept) | 5.92 | $4.63-7.55$ | $<\mathbf{0 . 0 0 1}$ |
| Time | 4.13 | $3.98-4.30$ | $<\mathbf{0 . 0 0 1}$ |
| Gender | 1.07 | $0.99-1.16$ | 0.071 |
| SPED | 0.54 | $0.48-0.61$ | $<\mathbf{0 . 0 0 1}$ |
| ELL | 0.84 | $0.73-0.97$ | $\mathbf{0 . 0 1 4}$ |
| group | 1.11 | $0.96-1.28$ | 0.173 |
| Tigr | 0.98 | $0.95-1.00$ | $\mathbf{0 . 0 4 4}$ |

Random Effects

| $\sigma^{2}$ | 0.02 |
| :--- | :--- |
| $\tau_{00}$ student_id:class_name | 0.39 |
| $\tau_{00}$ class_name | 0.05 |
| ICC | 0.97 |
| $\mathrm{~N}_{\text {student_id }}$ | 1095 |
| $\mathrm{~N}_{\text {class_name }}$ | 59 |
| Observations | 2190 |
| Marginal $\mathrm{R}^{2}$ / Conditional $\mathrm{R}^{2}$ | $0.534 / 0.984$ |

## Kindergarten: Above/Below Benchmark Comparisons

Note: The first column contains data for students Above Benchmark, the second column contains data for students Below Benchmark. The variable of interest is "Tigr", which represents the interaction between "Time" and "Group" and tells us whether the amount growth on the outcome measure is different for students in the control versus treatment groups.

| Predictors | ark 6 comp 1 cap |  |  | ark 6 comp 2 cap |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Estimates | CI | $p$ | Estimates | CI | $p$ |
| (Intercept) | 79.94 | 62.84-97.03 | $<0.001$ | 136.77 | $123.10-150.44$ | $<0.001$ |
| Time | 28.83 | $20.33-37.33$ | $<0.001$ | 21.87 | $15.55-28.18$ | $<0.001$ |
| Gender | -3.58 | -9.07-1.90 | 0.200 | -1.01 | -6.07-4.06 | 0.696 |
| SPED | -43.71 | $-52.78--34.64$ | <0.001 | -25.54 | $-34.78-16.30$ | <0.001 |
| ELL | -13.09 | -23.39--2.79 | 0.013 | 3.64 | $-5.55-12.84$ | 0.438 |
| group | -3.42 | -13.93-7.09 | 0.524 | 7.00 | -1.21-15.21 | 0.095 |
| Tigr | 2.57 | $-2.74-7.89$ | 0.343 | -2.14 | -6.02-1.74 | 0.280 |
| Random Effects |  |  |  |  |  |  |
| $\sigma^{2}$ | 989.67 |  |  | 466.76 |  |  |
| $\tau_{00}$ | 1541.07 student_id.class_name |  |  | 1319.37 student_id.class_name |  |  |
|  | 67.67 class_name |  |  | 31.13 class_name |  |  |
| ICC | 0.62 |  |  | 0.74 |  |  |
| N | 1079 student_id |  |  | 963 student_id |  |  |
|  | 59 class_name |  |  | 60 class_name |  |  |
| Observations | 2158 |  |  | 1926 |  |  |
| Marginal $\mathrm{R}^{2}$ / Conditional $\mathrm{R}^{2}$ | $0.147 / 0.675$ |  |  | $0.071 / 0.761$ |  |  |

Kindergarten: Above/Below Benchmark Comparisons (Well Below Benchmark students excluded)

| Predictors | ark 6 comp K |  |  |
| :---: | :---: | :---: | :---: |
|  | Incidence Rate Ratios | CI | $p$ |
| (Intercept) | 3.54 | $2.88-4.34$ | $<0.001$ |
| Time | 5.49 | $4.93-6.11$ | $<0.001$ |
| Gender | 1.01 | 0.99-1.04 | 0.277 |
| SPED | 0.99 | 0.96-1.02 | 0.419 |
| ELL | 1.01 | 0.96-1.05 | 0.802 |
| group | 1.00 | 0.88-1.13 | 0.940 |
| Tigr | 1.00 | 0.93-1.07 | 0.932 |
| Random Effects |  |  |  |
| $\sigma^{2}$ | 0.02 |  |  |
| $\tau_{00}$ student_id:class_name | 0.00 |  |  |
| $\tau_{00}$ class_name | 0.00 |  |  |
| $\mathrm{N}_{\text {student_id }}$ | 375 |  |  |
| $\mathrm{N}_{\text {class_name }}$ | 59 |  |  |
| Observations | 439 |  |  |
| Marginal $\mathrm{R}^{2} /$ Conditional $\mathrm{R}^{2}$ | 0.974 / NA |  |  |

Kindergarten: Teacher Training, Experimental Group

|  | ark 6 comp K |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Incidence Rate Ratios | $C I$ | $p$ |
| (Intercept) | 6.69 | $5.97-7.51$ | $<\mathbf{0 . 0 0 1}$ |
| Time | 3.98 | $3.91-4.05$ | $<\mathbf{0 . 0 0 1}$ |
| Gender | 1.14 | $1.03-1.27$ | $\mathbf{0 . 0 0 9}$ |
| SPED | 0.58 | $0.50-0.68$ | $<\mathbf{0 . 0 0 1}$ |
| ELL | 0.86 | $0.70-1.04$ | 0.128 |
| sort | 1.23 | $0.98-1.55$ | 0.068 |
| sortxtime | 0.94 | $0.91-0.98$ | $\mathbf{0 . 0 0 5}$ |

## Random Effects

| $\sigma^{2}$ | 0.01 |
| :--- | :--- |
| $\tau_{00 \text { student_id:class_name }}$ | 0.37 |
| $\tau_{00 \text { class_name }}$ | 0.04 |
| ICC | 0.97 |
| $\mathrm{~N}_{\text {student_id }}$ | 598 |
| $\mathrm{~N}_{\text {class_name }}$ | 32 |
| Observations | 1196 |

Marginal $\mathrm{R}^{2}$ / Conditional $\mathrm{R}^{2} 0.547 / 0.984$

|  | ark 6 comp K |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Incidence Rate Ratios | $C I$ | $p$ |
| (Intercept) | 6.54 | $5.69-7.52$ | $<\mathbf{0 . 0 0 1}$ |
| Time | 4.01 | $3.93-4.09$ | $<\mathbf{0 . 0 0 1}$ |
| Gender | 1.00 | $0.89-1.12$ | 0.944 |
| SPED | 0.50 | $0.42-0.60$ | $<\mathbf{0 . 0 0 1}$ |
| ELL | 0.83 | $0.68-1.01$ | 0.064 |
| sort | 1.32 | $0.99-1.75$ | 0.057 |
| sortxtime | 1.03 | $0.99-1.08$ | 0.184 |

## Random Effects

| $\sigma^{2}$ | 0.02 |
| :--- | :--- |
| $\tau_{00}$ student_id:class_name | 0.40 |
| $\tau_{00}$ class_name | 0.06 |
| ICC | 0.97 |
| $\mathrm{~N}_{\text {student_id }}$ | 497 |
| $\mathrm{~N}_{\text {class_name }}$ | 27 |

Observations ..... 994
Marginal $\mathrm{R}^{2} /$ Conditional $\mathrm{R}^{2} \quad 0.538 / 0.985$

Kindergarten: First Sound Fluency (FSF)

## fsf k

| Predictors | Incidence Rate Ratios | CI | $p$ |
| :---: | :---: | :---: | :---: |
| (Intercept) | 2.89 | 2.16-3.88 | <0.001 |
| Time | 2.85 | 2.66-3.05 | <0.001 |
| Gender | 1.13 | $1.03-1.24$ | 0.008 |
| SPED | 0.41 | 0.36-0.48 | <0.001 |
| ELL | 0.76 | 0.64-0.90 | 0.002 |
| group | 1.18 | 0.99-1.41 | 0.068 |
| Tigr | 0.96 | 0.92-1.00 | 0.032 |
| Random Effects |  |  |  |
| $\sigma^{2}$ | 0.06 |  |  |
| $\tau_{00}$ student_id:class_name | 0.55 |  |  |
| $\tau_{00}$ class_name | 0.07 |  |  |
| ICC | 0.92 |  |  |
| $\mathrm{N}_{\text {student_id }}$ | 1095 |  |  |
| $\mathrm{N}_{\text {class_name }}$ | 59 |  |  |
| Observations | 2190 |  |  |
| Marginal $\mathrm{R}^{2}$ / Conditional $\mathrm{R}^{2}$ | 0.334 / 0.944 |  |  |

Kindergarten: Letter Naming Fluency (LNF)

|  | $\operatorname{lnf} \mathbf{k}$ |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Incidence Rate Ratios | $C I$ | $p$ |
| (Intercept) | 7.63 | $6.14-9.47$ | $<\mathbf{0 . 0 0 1}$ |
| Time | 2.07 | $1.96-2.19$ | $<\mathbf{0 . 0 0 1}$ |
| Gender | 1.01 | $0.94-1.08$ | 0.865 |
| SPED | 0.69 | $0.62-0.77$ | $<\mathbf{0 . 0 0 1}$ |
| ELL | 0.86 | $0.75-0.98$ | $\mathbf{0 . 0 2 4}$ |
| group | 1.10 | $0.97-1.26$ | 0.142 |
| Tigr | 0.97 | $0.93-1.00$ | $\mathbf{0 . 0 3 9}$ |

## Random Effects

| $\sigma^{2}$ | 0.04 |
| :--- | :--- |
| $\tau_{00}$ student_id:class_name | 0.34 |
| $\tau_{00}$ class_name | 0.03 |
| ICC | 0.90 |
| $\mathrm{~N}_{\text {student_id }}$ | 1095 |
| $\mathrm{~N}_{\text {class_name }}$ | 59 |

Observations 2190
Marginal $\mathrm{R}^{2}$ / Conditional $\mathrm{R}^{2} \quad 0.240$ / 0.926

## Appendix 2: First Grade Results

- Composite score: $(\mathrm{B}=4.62, \mathrm{p}=0.30)$ - no significant differences between treatment and control groups
- Nonsense Word Fluency Correct Letter-Sounds (CLS) score: ( $\mathrm{B}=2.50, \mathbf{p}=\mathbf{0 . 0 4 7}$ ) - significant differences between treatment and control groups
- Nonsense Word Fluency Whole Words Read (WWR) score: ( $B=1.38, \mathbf{p}=\mathbf{0 . 0 1 9}$ ) - significant differences between treatment and control groups

For Below or Well Below Benchmark students:

- Composite score: $(\mathrm{B}=1.01, \mathrm{p}=0.68)$ - no significant differences between treatment and control group s


## For Below Benchmark students:

- Composite score: $(\mathrm{B}=-.09, \mathrm{p}=0.949)$ - no significant differences between treatment and control groups


## Exploring teacher training:

- EXPERIMENTAL GROUP: Composite score: ( $\mathrm{B}=-14.06, \mathrm{p}=0.025$ ) - significant effect of teacher training
- CONTROL GROUP: Composite score: ( $\mathrm{B}=-7.16, \mathrm{p}=0.267$ ) - significant effect of teacher training

First Grade: Composite Score

|  | ark 6 comp 1 |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | 60.73 | $34.15-87.30$ | $<0.001$ |
| Time | 50.39 | $36.30-64.49$ | $<0.001$ |
| Gender | -6.19 | $-13.90-1.52$ | 0.115 |
| SPED | -50.58 | $-63.33--37.82$ | $<0.001$ |
| ELL | -15.80 | $-30.29--1.32$ | 0.033 |
| group | -5.34 | $-21.73-11.05$ | 0.523 |
| Tigr | 4.62 | $-4.19-13.43$ | 0.304 |

Random Effects

| $\sigma^{2}$ | 2716.55 |
| :--- | :--- |
| $\tau_{00}$ student_id:class_name | 2659.82 |
| $\tau_{00 \text { class_name }}$ | 138.44 |
| ICC | 0.51 |
| $\mathrm{~N}_{\text {student_id }}$ | 1079 |
| $\mathrm{~N}_{\text {class_name }}$ | 59 |

Observations 2158

Marginal $\mathrm{R}^{2} /$ Conditional $\mathrm{R}^{2} \quad 0.162 / 0.587$

## First Grade: Above/Below Benchmark Comparisons

Note: The first column contains data for students Above Benchmark, the second column contains data for students Below Benchmark. The variable of interest is "Tigr", which represents the interaction between "Time" and "Group" and tells us whether growth in the outcome is different for students in the control versus treatment groups.

| Predictors | ark 6 comp 1 |  |  | ark 6 comp 1 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Estimates | CI | $p$ | Estimates | CI | $p$ |
| (Intercept) | 31.63 | -6.12-69.38 | 0.100 | 87.48 | $74.05-100.91$ | $<0.001$ |
| Time | 99.00 | 78.51-119.48 | $<0.001$ | -4.88 | $-12.61-2.85$ | 0.216 |
| Gender | -6.44 | $-15.41-2.54$ | 0.160 | 2.13 | $-1.63-5.89$ | 0.267 |
| SPED | -5.30 | $-24.70-14.10$ | 0.592 | -27.61 | -33.16--22.07 | $<0.001$ |
| ELL | -4.25 | $-22.09-13.59$ | 0.641 | -6.01 | $-12.77-0.75$ | 0.081 |
| group | 3.41 | $-20.27-27.08$ | 0.778 | -1.11 | -9.35-7.14 | 0.792 |
| Tigr | -0.89 | $-13.77-11.98$ | 0.892 | 1.01 | $-3.84-5.86$ | 0.683 |
| Random Effects |  |  |  |  |  |  |
| $\sigma^{2}$ | 2108.20 |  |  | 383.73 |  |  |
| $\tau_{00}$ | 1627.02 student_id:class_name |  |  | 403.93 student_id:class_name |  |  |
|  | 129.81 class_name |  |  | 37.16 class_name |  |  |
| ICC | 0.45 |  |  | 0.53 |  |  |
| N | 635 student_id |  |  | 753 student_id |  |  |
|  | 59 class_name |  |  | $59 \text { class_name }$ |  |  |
| Observations | 961 |  |  | 1197 |  |  |
| Marginal $\mathrm{R}^{2} /$ Conditional $\mathrm{R}^{2}$ | $0.370 / 0.656$ |  |  | $0.110 / 0.586$ |  |  |

First Grade: Above/Below Benchmark Comparisons (Well Below Benchmark students excluded)

|  | ark 6 comp 1 |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | 92.48 | $85.85-99.11$ | $<0.001$ |
| Time | 9.91 | $5.52-14.31$ | $<0.001$ |
| Gender | 1.35 | $-0.05-2.76$ | 0.058 |
| SPED | 0.35 | $-2.57-3.28$ | 0.814 |
| ELL | 1.17 | $-1.47-3.80$ | 0.385 |
| group | 0.65 | $-3.43-4.72$ | 0.755 |
| Tigr | -0.09 | $-2.84-2.66$ | 0.949 |
| Random Effects |  |  |  |
| $\sigma^{2}$ | 36.29 |  |  |
| $\tau_{00}$ student_id:class_name | 5.93 |  |  |
| $\tau_{00}$ class_name | 0.00 |  |  |
| $\mathrm{~N}_{\text {student_id }}$ | 317 |  |  |
| $\mathrm{~N}_{\text {class_name }}$ | 59 |  |  |
| Observations | 353 |  |  |
| Marginal $^{2} /$ Conditional $\mathrm{R}^{2}$ | $0.391 / \mathrm{NA}$ |  |  |

First Grade: Teacher Training, Experimental Group

|  | ark 6 comp 1 |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | 45.85 | $28.59-63.10$ | $<\mathbf{0 . 0 0 1}$ |
| Time | 65.92 | $57.67-74.17$ | $<\mathbf{0 . 0 0 1}$ |
| Gender | -3.65 | $-14.39-7.08$ | 0.505 |
| SPED | -50.28 | $-68.18--32.38$ | $<\mathbf{0 . 0 0 1}$ |
| ELL | -11.92 | $-32.06-8.21$ | 0.246 |
| sort | 5.61 | $-18.34-29.57$ | 0.646 |
| sortxtime | -14.06 | $-26.39--1.72$ | $\mathbf{0 . 0 2 5}$ |

## Random Effects

| $\sigma^{2}$ | 2746.57 |
| :--- | :--- |
| $\tau_{00}$ student_id:class_name | 2653.18 |
| $\tau_{00}$ class_name | 221.95 |
| ICC | 0.51 |
| $\mathrm{~N}_{\text {student_id }}$ | 561 |
| $\mathrm{~N}_{\text {class_name }}$ | 30 |
| Observations | 1122 |

Marginal $R^{2}$ / Conditional $R^{2} \quad 0.174 / 0.597$

First Grade: Teacher Training, Control Group

|  | ark 6 comp 1 |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | 58.39 | $40.92-75.86$ | $<\mathbf{0 . 0 0 1}$ |
| Time | 51.04 | $41.63-60.45$ | $<\mathbf{0 . 0 0 1}$ |
| Gender | -9.53 | $-20.56-1.51$ | 0.091 |
| SPED | -49.46 | $-67.49--31.42$ | $<\mathbf{0 . 0 0 1}$ |
| ELL | -25.44 | $-45.90-4.98$ | $\mathbf{0 . 0 1 5}$ |
| sort | -1.47 | $-23.37-20.44$ | 0.896 |
| sortxtime | 7.16 | $-5.48-19.80$ | 0.267 |

## Random Effects

| $\sigma^{2}$ | 2661.62 |
| :--- | :--- |
| $\tau_{00}$ student_id:class_name | 2670.00 |
| $\tau_{00}$ class_name | 0.00 |
| ICC | 0.50 |
| $\mathrm{~N}_{\text {student_id }}$ | 518 |
| $\mathrm{~N}_{\text {class_name }}$ | 29 |
| Observations | 1036 |

Marginal $\mathrm{R}^{2} /$ Conditional $\mathrm{R}^{2} \quad 0.165 / 0.583$

First Grade: Nonsense Word Fluency Correct Letter Sounds (CLS)

|  | cls l |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | 16.29 | $7.52-25.07$ | $<0.001$ |
| Time | 17.83 | $13.87-21.79$ | $<0.001$ |
| Gender | -3.81 | $-6.71--0.91$ | $\mathbf{0 . 0 1 0}$ |
| SPED | -16.43 | $-21.24--11.63$ | $<0.001$ |
| ELL | -2.53 | $-8.00-2.94$ | 0.364 |
| group | -3.11 | $-8.50-2.28$ | 0.259 |
| Tigr | 2.50 | $0.03-4.98$ | $\mathbf{0 . 0 4 7}$ |

## Random Effects

| $\sigma^{2}$ | 214.35 |
| :--- | :--- |
| $\tau_{00}$ student_id:class_name | 460.10 |
| $\tau_{00}$ class_name | 27.22 |
| ICC | 0.69 |
| $\mathrm{~N}_{\text {student_id }}$ | 1079 |
| $\mathrm{~N}_{\text {class_name }}$ | 59 |
| Observations | 2158 |
| Marginal $\mathrm{R}^{2} /$ Conditional $\mathrm{R}^{2}$ | $0.170 / 0.746$ |

First Grade: Nonsense Word Fluency Whole Words Read (WWR)

|  | wwr l |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | -0.25 | $-3.93-3.43$ | 0.894 |
| Time | 6.00 | $4.16-7.84$ | $<\mathbf{0 . 0 0 1}$ |
| Gender | -1.31 | $-2.39--0.23$ | $\mathbf{0 . 0 1 7}$ |
| SPED | -4.72 | $-6.51--2.93$ | $<0.001$ |
| ELL | -1.02 | $-3.06-1.01$ | 0.325 |
| group | -1.31 | $-3.58-0.96$ | 0.259 |
| Tigr | 1.38 | $0.23-2.53$ | 0.019 |
| Random Effects |  |  |  |
| $\sigma^{2}$ | 46.36 |  |  |
| $\tau_{00}$ student_id:class_name | 55.41 |  |  |
| $\tau_{00}$ class_name | 4.02 |  |  |
| ICC $^{N_{\text {student_id }}}$ | 0.56 |  |  |
| $\mathrm{~N}_{\text {class_name }}$ | 1079 |  |  |
| Observations | 59 |  |  |
| Marginal R $^{2} /$ Conditional $R^{2}$ | $0.151 / 0.628$ |  |  |

## Appendix 3: Second Grade Results

- Composite score: $(\mathrm{B}=7.17, \mathbf{p}=\mathbf{0 . 0 2 4})$ - significant differences between treatment and control groups
- Oral Reading Fluency Words Correct Per Minute (ORF) score: ( $\mathrm{B}=3.93$, $\mathbf{p}<\mathbf{0 . 0 0 1}$ ) significant differences between treatment and control groups
- Oral Reading Fluency Error (ERR) score: ( $\mathrm{B}=0.41, \mathrm{p}=0.18$ ) - no significant differences between treatment and control groups
- Retell Total (RETELL) score: $(\mathrm{B}=-0.03, \mathrm{p}=0.98)$ - no significant differences between treatment and control groups
- Retell Quality of Response (RETELL QR) score: $(B=-0.07, \mathrm{p}=0.29)$ - no significant differences between treatment and control groups

For Below or Well Below Benchmark students:

- Composite score: $(\mathrm{B}=-3.39, \mathrm{p}=0.51)$ - no significant differences between treatment and control groups

For Below Benchmark students:

- Composite score: ( $\mathrm{B}=0.61, \mathrm{p}=0.834$ ) - no significant differences between treatment and control groups


## Exploring teacher training:

- EXPERIMENTAL GROUP: Composite score: $(\mathrm{B}=-1.23, \mathrm{p}=0.784)$ - no significant effect of teacher training
- CONTROL GROUP: Composite score: ( $\mathrm{B}=3.25, \mathrm{p}=0.474)$ - no significant effect of teacher training

Second Grade: Composite Score

|  | ark 6 comp 2 |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | 122.73 | $98.63-146.84$ | $<0.001$ |
| Time | 46.74 | $36.63-56.85$ | $<\mathbf{0 . 0 0 1}$ |
| Gender | 3.68 | $-5.07-12.42$ | 0.410 |
| SPED | -39.87 | $-55.86--23.89$ | $<\mathbf{0 . 0 0 1}$ |
| ELL | 16.54 | $0.58-32.50$ | $\mathbf{0 . 0 4 2}$ |
| group | 1.65 | $-12.84-16.15$ | 0.823 |
| Tigr | 7.17 | $0.96-13.39$ | $\mathbf{0 . 0 2 4}$ |
| Random Effects |  |  |  |
| $\sigma^{2}$ | 1196.62 |  |  |
| $\tau_{00}$ student_id:class_name | 4007.74 |  |  |
| $\tau_{00}$ class_name | 180.01 |  |  |
| ICC | 0.78 |  |  |
| $\mathrm{~N}_{\text {student_id }}$ | 963 |  |  |
| $\mathrm{~N}_{\text {class_name }}$ | 60 |  |  |
| Observations | 1926 |  |  |
| Marginal $^{2} /$ Conditional $\mathrm{R}^{2}$ | $0.160 / 0.813$ |  |  |

## Second Grade: Above/Below Benchmark Comparisons

Note: The first column contains data for students Above Benchmark, the second column contains data for students Below Benchmark. The variable of interest is "Tigr", which represents the interaction between "Time" and "Group" and tells us whether growth in the outcome is different for students in the control versus treatment groups.

| Predictors | ark 6 comp 2 |  |  | ark 6 comp 2 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Estimates | CI | $p$ | Estimates | CI | $p$ |
| (Intercept) | 147.02 | 126.69-167.36 | $<0.001$ | 76.63 | 48.44-104.83 | $<0.001$ |
| Time | 52.75 | 41.98-63.53 | $<0.001$ | 32.96 | $17.12-48.80$ | $<0.001$ |
| Gender | 6.24 | 0.17-12.32 | 0.044 | 4.42 | -4.38-13.21 | 0.324 |
| SPED | -6.75 | -19.15-5.65 | 0.286 | -24.49 | -37.66--11.32 | $<0.001$ |
| ELL | 15.17 | 4.33-26.01 | 0.006 | -2.29 | -19.03-14.46 | 0.789 |
| group | -2.62 | -14.78-9.54 | 0.673 | -1.59 | -19.29-16.11 | 0.860 |
| Tigr | 6.86 | 0.33-13.39 | 0.039 | -3.39 | -13.58-6.79 | 0.513 |
| Random Effects |  |  |  |  |  |  |
| $\sigma^{2}$ | 926.29 |  |  | 629.75 |  |  |
| $\tau_{00}$ | 1327.01 student_id:class_name |  |  | 983.00 student_id:class_name |  |  |
|  | 48.05 class_name |  |  | 64.21 class_name |  |  |
| ICC | 0.60 |  |  | 0.62 |  |  |
| N | 802 student_id |  |  | $309{ }_{\text {student_id }}$ |  |  |
|  | $60 \text { class_name }$ |  |  | $60 \text { class_name }$ |  |  |
| Observations | 1456 |  |  | 470 |  |  |
| Marginal $\mathrm{R}^{2} /$ Conditional $\mathrm{R}^{2}$ | $0.315 / 0.724$ |  |  | $0.146 / 0.679$ |  |  |

Second Grade: Above/Below Benchmark Comparisons (Well Below Benchmark students excluded)

|  | ark 6 comp 2 |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | 85.36 | $71.61-99.11$ | $<0.001$ |
| Time | 42.58 | $33.94-51.22$ | $<0.001$ |
| Gender | -1.68 | $-4.92-1.55$ | 0.306 |
| SPED | -6.46 | $-12.43--0.48$ | 0.034 |
| ELL | -0.14 | $-6.21-5.94$ | 0.965 |
| group | -1.12 | $-10.00-7.75$ | 0.803 |
| Tigr | 0.61 | $-5.11-6.33$ | 0.834 |

Random Effects

| $\sigma^{2}$ | 85.01 |
| :--- | :--- |
| $\tau_{00 \text { student_id:class_name }}$ | 39.16 |
| $\tau_{00 \text { class_name }}$ | 0.00 |
| $\mathrm{~N}_{\text {student_id }}$ | 183 |
| $\mathrm{~N}_{\text {class_name }}$ | 56 |
| Observations | 216 |
| Marginal $\mathrm{R}^{2} /$ Conditional $\mathrm{R}^{2}$ | $0.848 / \mathrm{NA}$ |

Second Grade: Teacher Training, Experimental Group
ark 6 comp 2

| Predictors | Estimates | $C I$ | $p$ |
| :--- | :---: | :---: | :---: |
| (Intercept) | 122.38 | $107.37-137.39$ | $<0.001$ |
| Time | 61.60 | $55.91-67.29$ | $<0.001$ |
| Gender | 8.04 | $-3.74-19.82$ | 0.181 |
| SPED | -42.81 | $-64.21--21.42$ | $<0.001$ |
| ELL | 23.35 | $-1.99-48.68$ | 0.071 |
| sort | 2.87 | $-17.82-23.56$ | 0.786 |
| sortxtime | -1.23 | $-10.00-7.54$ | 0.784 |

## Random Effects

| $\sigma^{2}$ | 1290.96 |
| :--- | :--- |
| $\tau_{00}$ student_id.class_name | 3968.36 |
| $\tau_{00}$ class_name | 242.20 |
| ICC | 0.77 |
| $\mathrm{~N}_{\text {student_id }}$ | 529 |
| $\mathrm{~N}_{\text {class_name }}$ | 33 |
| Observations | 1058 |

Marginal $\mathrm{R}^{2} /$ Conditional $\mathrm{R}^{2} \quad 0.169 / 0.805$

Second Grade: Teacher Training, Control Group

|  | ark 6 comp 2 |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | 132.38 | $117.25-147.50$ | $<\mathbf{0 . 0 0 1}$ |
| Time | 52.56 | $46.83-58.29$ | $<\mathbf{0 . 0 0 1}$ |
| Gender | -1.39 | $-14.49-11.72$ | 0.836 |
| SPED | -36.55 | $-60.75--12.36$ | $\mathbf{0 . 0 0 3}$ |
| ELL | 11.82 | $-8.81-32.45$ | 0.261 |
| sort | -12.38 | $-32.81-8.05$ | 0.235 |
| sortxtime | 3.25 | $-5.65-12.15$ | 0.474 |
| Random Effects |  |  |  |
| $\sigma^{2}$ | 1085.55 |  |  |
| $\tau_{00}$ student_id:class_name | 4071.56 |  |  |
| $\tau_{00}$ class_name | 115.15 |  |  |
| ICC | 0.79 |  |  |
| $\mathrm{~N}_{\text {student_id }}$ | 434 |  |  |
| $\mathrm{~N}_{\text {class_name }}$ | 27 |  |  |
| Observations | 868 |  |  |
| ${\text { Marginal } \mathrm{R}^{2} / \text { Conditional } \mathrm{R}^{2}}$ $0.141 / 0.823$  |  |  |  |

Second Grade: Oral Reading Fluency Words Correct Per Minute (ORF)

|  | orf 2 |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| Intercept) | 48.07 | $38.52-57.62$ | $<0.001$ |
| Time | 12.99 | $9.62-16.35$ | $<0.001$ |
| Gender | 1.52 | $-2.36-5.40$ | 0.442 |
| SPED | -10.39 | $-17.49--3.30$ | 0.004 |
| ELL | 9.90 | $2.83-16.98$ | 0.006 |
| group | 1.20 | $-4.50-6.91$ | 0.679 |
| Tigr | 3.93 | $1.86-6.00$ | $<0.001$ |
| Random Effects |  |  |  |
| $\sigma^{2}$ | 132.65 |  |  |
| $\tau_{00}$ student_id:class_name | 842.37 |  |  |
| $\tau_{00}$ class_name | 30.44 |  |  |
| ICC | 0.87 |  |  |
| $\mathrm{~N}_{\text {student_id }}$ | 963 |  |  |
| $\mathrm{~N}_{\text {class_name }}$ | 60 |  |  |
| Observations | 1926 |  |  |
| Marginal $^{2} /$ Conditional $\mathrm{R}^{2}$ | $0.106 / 0.882$ |  |  |

Second Grade: Oral Reading Fluency Errors (ERR)

|  | err 2 |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | 6.78 | $5.08-8.48$ | $<0.001$ |
| Time | -2.36 | $-3.34--1.38$ | $<0.001$ |
| Gender | 0.12 | $-0.30-0.55$ | 0.566 |
| SPED | 1.73 | $0.95-2.50$ | $<0.001$ |
| ELL | -0.09 | $-0.86-0.69$ | 0.824 |
| group | -0.61 | $-1.64-0.43$ | 0.249 |
| Tigr | 0.41 | $-0.19-1.01$ | 0.182 |

Random Effects

| $\sigma^{2}$ | 11.23 |
| :--- | :--- |
| $\tau_{00}$ student_id:class_name | 5.32 |

$\tau_{00}$ class_name $\quad 0.30$
ICC 0.33
$\mathrm{N}_{\text {student_id }} \quad 963$

| $\mathrm{N}_{\text {class_name }}$ | 60 |
| :--- | :--- |
| Observations | 1926 |

Marginal $\mathrm{R}^{2}$ / Conditional $\mathrm{R}^{2} \quad 0.055 / 0.370$

Second Grade: Retell Total (RETELL)

|  | retell 2 |  |  |
| :--- | :---: | :---: | :---: |
| Predictors | Estimates | $C I$ | $p$ |
| (Intercept) | 13.44 | $7.91-18.97$ | $<0.001$ |
| Time | 5.84 | $2.94-8.74$ | $<0.001$ |
| Gender | 3.31 | $1.72-4.90$ | $<0.001$ |
| SPED | -7.48 | $-10.39-4.57$ | $<0.001$ |
| ELL | 1.31 | $-1.60-4.21$ | 0.377 |
| group | 2.33 | $-1.02-5.68$ | 0.173 |
| Tigr | -0.03 | $-1.81-1.76$ | 0.978 |

Random Effects

| $\sigma^{2}$ | 98.67 |
| :--- | :--- |
| $\tau_{00}$ student_id:class_name | 103.17 |
| $\tau_{00}$ class_name | 5.92 |
| ICC | 0.53 |
| $\mathrm{~N}_{\text {student_id }}$ | 963 |
| $\mathrm{~N}_{\text {class_name }}$ | 60 |
| Observations | 1926 |

Marginal $\mathrm{R}^{2} /$ Conditional $\mathrm{R}^{2} \quad 0.079 / 0.563$

Second Grade: Retell Quality of Response (RETELLQR)

| Predictors | retellqr 2 |  |  |
| :---: | :---: | :---: | :---: |
|  | Estimates | CI | $p$ |
| (Intercept) | 1.39 | 0.97-1.80 | $<0.001$ |
| Time | 0.49 | $0.27-0.71$ | $<0.001$ |
| Gender | 0.16 | 0.06-0.26 | 0.002 |
| SPED | -0.47 | -0.65--0.28 | $<0.001$ |
| ELL | 0.07 | $-0.12-0.26$ | 0.466 |
| group | 0.13 | $-0.12-0.38$ | 0.320 |
| Tigr | -0.07 | -0.21-0.06 | 0.288 |
| Random Effects |  |  |  |
| $\sigma^{2}$ | 0.57 |  |  |
| $\tau_{00}$ student_id:class_name | 0.35 |  |  |
| $\tau_{00}$ class_name | 0.04 |  |  |
| ICC | 0.41 |  |  |
| $\mathrm{N}_{\text {student_id }}$ | 963 |  |  |
| $\mathrm{N}_{\text {class_name }}$ | 60 |  |  |
| Observations | 1926 |  |  |
| Marginal $\mathrm{R}^{2}$ / Conditional $\mathrm{R}^{2}$ | 0.061 / 0 | 443 |  |


[^0]:    ${ }^{1}$ As of Fall 2020, all students receive free lunch, so documentation on free-lunch status is not available at the student level.

[^1]:    ${ }^{2}$ Four students in first grade were tested Monday and Tuesday of the following week (September 20 and 21), due to absences.

[^2]:    ${ }^{3}$ As of Fall 2020, all students receive free lunch, so documentation on free-lunch status is not available at the student level.

