



95 Phonological Awareness Lessons Research Study

Kindergarten Results, Fall 2021

Prepared by Rachel Schechter, LXD Research

Analysis conducted by Lynch Research Associates

Introduction	3
Study Program Description	3
Comparison Programs	4
Table 1. ESSA-Level Evidence on Comparison School programs for Kindergarten	4
Fall Research Description	4
Reading Assessments	5
Table 2. Acadience Reading Subtests, Skill Coverage, and Benchmark Goals in Kindergarten	5
Student and School Demographics	6
Fall Implementation	6
95 Percent Group Coaching Summary	6
PASI and PSI Implementation Description	6
How many students have received PA Lessons?	6
Table 3. Number of Students in Each Cycle by Study	7
Are students progressing in the program as expected?	7
Figure 1. Percent of Students per Skill by Cycle	7
Results	8
Sample Descriptions	8
Table 4. Sample sizes at baseline and midyear by study and treatment group	8
Table 5. Sample descriptive for treatment and control group by study	9
Analytical Approach	9

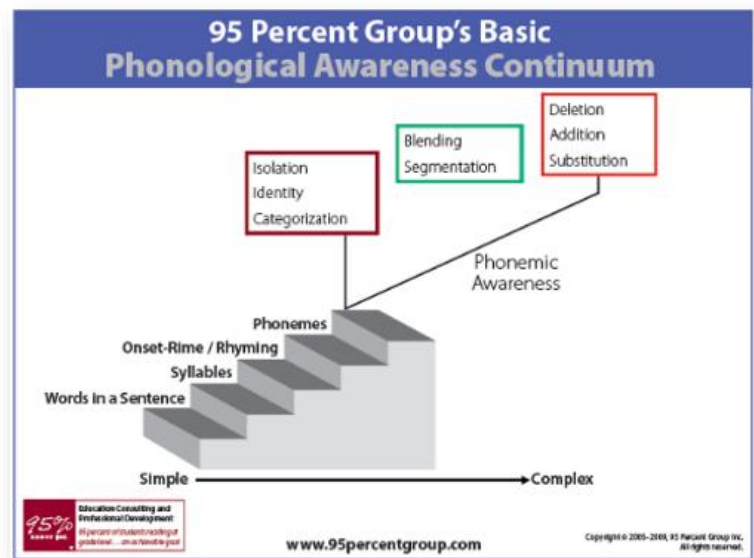
California Study MOY-BOY Statistical Results	10
All Students	10
Table 6. California Results for All Students	10
Student BOY Groups: At/Above Benchmark and Below/Well Below Benchmark	10
Table 7. Composite Score Results, Differences by BOY Benchmark Group	11
Utah Study MOY-BOY Results	12
Table 8. Utah Results for All Students	12
Table 9. Utah Composite Score Results, Differences by BOY Benchmark Group	13
Key Findings	14
Treatment vs. Comparison Schools	14
Treatment Schools Only	15
Conclusion and Next Steps	16
Appendices	17
Appendix 1: California Results	17
<i>Composite Score</i>	18
<i>At or Above Benchmark Comparisons</i>	19
<i>Below Benchmark Comparisons</i>	19
<i>FSF</i>	21
<i>LNF</i>	22
Appendix 2: Utah Results	23
<i>Composite Score</i>	24
<i>At or Above Benchmark Comparisons</i>	25
<i>Below Benchmark Comparisons</i>	25

Introduction

Learning Experience Design Research (LXD Research) is an independent evaluation, research, and consulting division within Charles River Media Group focusing on education. LXD Research designs rigorous research studies, multifaceted data analytic reporting, and dynamic content to disseminate insights. For 95 Percent Group, LLC, the team is conducting two studies during 2021/22 of the impact and implementation of Phonological Awareness Lessons (PA Lessons) in Utah and California. The studies use quasi-experimental designs to generate evidence of the program's impact that aligns with evidence standards associated with ESSA. This interim report describes the Fall results from September 2021 to January 2022.

Study Program Description

The Phonological Awareness (PA) Lessons are designed primarily for Tier 2 or Tier 3 phonological awareness intervention in Grades and K. In the study, the PA Lesson intervention included the use of the an initial diagnostic screener, and then the use of the 95 Percent Group's Phonological Awareness Screener for Intervention™ (PASI) to group students into intervention groups based on skill needs every three weeks. Students who were Below Benchmark are identified for intervention through use of a curriculum-based measure or an early literacy screener assessment used by the district, and then placed into lessons along the Phonological Awareness Continuum through the PASI. The PA Lessons support students who are not meeting benchmarks through comprehensive lesson plans that target skills aligned with the Phonological Awareness Continuum, from readiness (understanding concepts and terms; applying language) through phonological awareness (syllables; onset rimes; phonemes). Students received 20-30 minutes of daily intervention through a push-in model, in small groups of three to four students who are at similar levels. Instructors monitor progress through alternate forms of the PASI and use this data to re-group students every 3 weeks based on the lowest skill on the continuum that needs support. Instruction is grounded in and aligned with evidence-based instructional practices in literacy. Once students reach mastery of skills for their grade level, they have completed the intervention.



Comparison Programs

In California, in the comparison schools survey, most teachers (42%) responded that they used their core curriculum, Wonders, to support Tier 2 and Tier 3 reading intervention. A small group of teachers also mentioned using Heggerty Phonemic Awareness resources (17%). In Utah, Tier 2 and Tier 3 PA and Phonics was covered by multiple products including but not limited to Imagine Learning (50%), Reading Horizons (50%), Heggerty (36%), Early Reading Intervention (36%), as well as the core curriculum Journeys (29%). Evidence related to these programs' effectiveness varies, with most not having any evidence with kindergartners (Table 1.). In both studies, comparison schools implemented intervention in a variety of ways, usually pulling students out for thirty minutes for Tier 3 and using small group instruction during the reading block for Tier 2.

Table 1. ESSA-Level Evidence on Comparison School programs for Kindergarten

Study	Product	Evidence for All Students	Evidence Tiers 2 - 3
California	Wonders	None	None
	Heggerty	None	None
Utah	Journeys	1 Strong Study	1 Strong Study
	Imagine Learning	None	None
	Reading Horizons Discovery®	None	None
	Early Reading Intervention	None	1 Strong Study (<350 students)

Fall Research Description

The goals of the Fall activities were to begin to understand the nature and extent of literacy program implementation in comparison schools, as well as to understand mid-year gains. During Fall 2021, district leaders supported data collection to initiate the study and support product coaching services. Those activities included:

- Conducting Acadience® Reading K-6 with all students at the beginning of the year (BOY)

- Conducting segments from the CORE Phonological Segmentation Test & CORE Phoneme Deletion Test (Consortium on Reaching Excellence in Education) with all students with Below Benchmark and Well Below Benchmark scores on Acadience
- Conducting PASI with treatment schools every three weeks for progress monitoring
- Sharing the results of the data collection activities with the research teams
- Conducting Acadience Reading with all students in the middle of the year (MOY)

This report focuses on the gains from BOY to MOY on Acadience Reading.

Reading Assessments

Acadience Reading K-6 assessments were administered by a special assessment team (not classroom teachers) at both school districts at the beginning of fall 2021. As a set of curriculum-based measures Acadience Reading assesses student development as a reader. Designed for universal screening and progress monitoring to determine the appropriate supports for each student, Acadience is administered three times per year in fall, winter, and spring. Assessments take between 3 and 11 minutes per student to complete. Scores include standardized scale scores and on-grade achievement-level placements. Kindergarten Acadience Reading subtests are listed in Table 2, along with the skills they assess and the benchmark goals for the times of year they are administered (the measures administered vary by time of year based on expected skill development). Note that the LNF measure does not have benchmark goals because it is an indicator of risk rather than an indicator of a basic early literacy skill. At each administration period, subtest scores are weighted and combined into a Composite Score, which is an overall indicator of reading ability.

Table 2. Acadience Reading Subtests, Skill Coverage, and Benchmark Goals in Kindergarten

Subtest	Indicators of These Basic Early Literacy Skills	Benchmark Goals		
		BOY	MOY	EOY
First Sound Fluency	Phonemic Awareness	X	X	
Phoneme Segmentation Fluency	Phonemic Awareness		X	X
Letter Naming Fluency	Indicator of Risk	N/A	N/A	N/A
Nonsense Word Fluency	The Alphabetic Principle and Basic Phonics		X	X
Composite	Overall Estimate of Reading Ability	X	X	X

[CORE Phonological Segmentation Test](#) & [CORE Phoneme Deletion Test](#) also administered by the special assessment team to both intervention and comparison students at the two participating districts in Fall 2021. The CORE surveys measure phonics and phonics-related skills in beginning reading. Each survey presents several lists of letters and words for students to identify or decode. Pseudowords are also included to assess decoding skills. The CORE Phonics Surveys were used as a diagnostic tool to place students into initial reading groups for this study and then the PASI tracked student progress from earlier skills to grade-level mastery (and then exiting intervention as appropriate). The surveys take approximately 10 minutes to administer.

Student and School Demographics

Student demographics that may be related to outcome measures were collected, including school, district, gender, grade, race/ethnicity, age, English language learner status, economic disadvantage status (the likely proxy is an indicator of whether a student qualifies for FRM), homeless status, migrant status, attendance rate, special education status, and whether or not a student was retained in a grade. School characteristics that may affect outcomes were also collected, including percent English language learners, percent students in special education, and total student enrollment.

Fall Implementation

95 Percent Group Coaching Summary

Training to support kindergarten teachers was provided in both school districts before school started. Coaches gave guidance on how to use the CORE assessments to place students in initial groups. The use of PASI and PSI began with Cycle 2, and the PA Lessons were used during intervention time. With each cycle, teachers created student groups to focus on specific PA or Phonics skills. Over time, students should advance through the 95 Percent Group PA Continuum. Coaches have been available to consult and discuss questions during the Fall and return for a visit in Winter 2022.

PASI and PSI Implementation Description

Teachers have been completing the Phonological Awareness Screening Inventory every three weeks as part of the intervention. The results of these screeners inform student groupings and the targeted skill for the cycle's lessons. This section of the report summarizes the number of students who have been identified and served by literacy intervention during cycles 2-5.

How many students have received PA Lessons?

Between 500 and 600 students have been grouped for intervention during each cycle between August and December 2021 (Table 3).



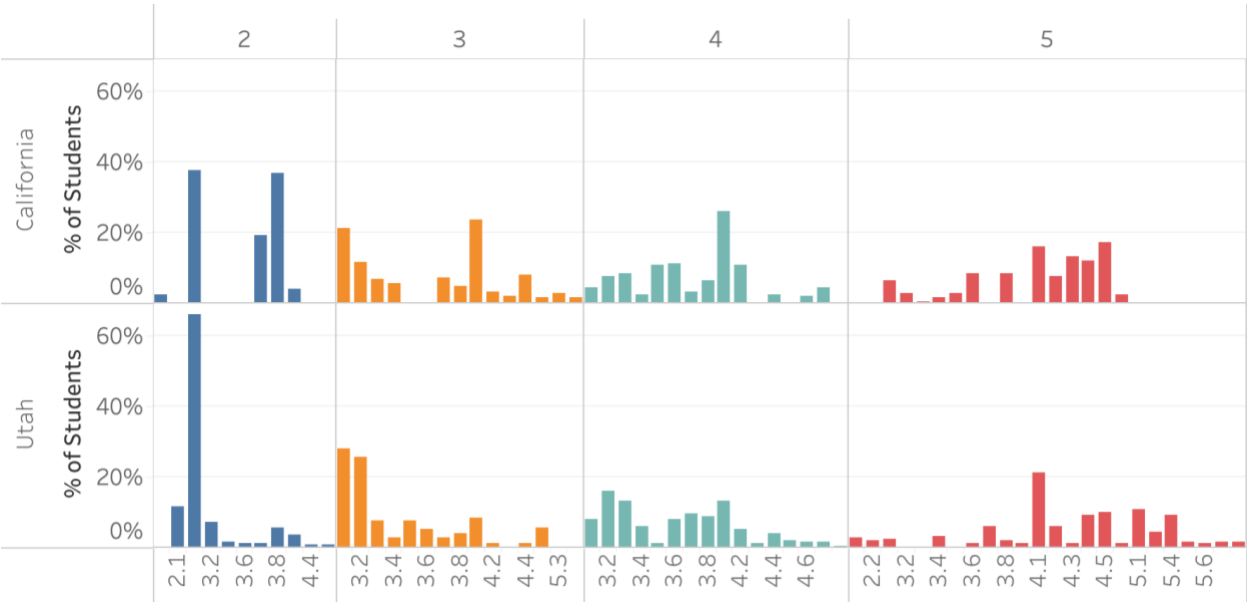
Table 3. Number of Students in Each Cycle by Study

	Cycle Number				
Location	2	3	4	5	Placed out by Cycle 5
California Study	218	190	158	168	50 (23%)
Utah Study	409	379	358	394	15 (4%)
Total	627	569	516	562	65 (10%)

Are students progressing in the program as expected?

Most students started the school year working on skills at the start of the skill progression for each program. As the cycles progressed, more students advanced through the programs. By January, kindergartners had completed 53% of the program. A larger percentage of students in Utah started with lower skill levels than in California, which seems to continue being ahead as the cycles continued. Some students have not progressed at the expected rate (Figure 1). Many of these students may be able to move up to more advanced lessons, even while they are working on the more basic skills.

Figure 1. Percent of Students per Skill by Cycle



Even though research has shown that PA lessons may have a diminishing effect when they are repeated too many times ([NRP 2000 Report](#)), many teachers struggle with allowing students to move on when they do not display consistent and complete mastery. PA is a difficult foundational skill for teachers to



teach without explicit and supportive training, for as expert readers their knowledge of spelling can interfere with their ability to break down the sounds (and teach others how to do it as well). Site visits with 95 Percent Group coaches that occurred since Cycle 5 will hopefully help teachers understand how to facilitate accelerated learning during the winter/spring months.

Results

Sample Descriptions

A total of 2,339 students had both BOY and MOY data across the two studies. Using a quasi-experimental design to examine the effects of the 95 Percent Group’s PA Lessons, a portion of schools used the walk-to-intervention program (treatment) and another portion did not (control). Of these students, 1,123 students were in the treatment group and 1,216 students were in the control group (see Table 4 for details).

In California, students in the treatment and control groups were similar regarding gender, English Language Learner status (ELL), Hispanic ethnicity, and rates of Foster/Homelessness.

In Utah, students in treatment and control were similar regarding gender and ELL status. However, Kindergarten students in the control group were more likely to receive special education services ($\chi^2=9.87$, $p = .002$; see Table 1a) and more likely to be White ($\chi^2=7.09$, $p = .008$). See Table 5 for details.

Table 4. Sample sizes at baseline and midyear by study and treatment group

Study	School Group	BOY	MOY	Sample
		# of Students	# of Students	# of Students
California	Treatment	455	476	443
	Control	567	593	502
	Total	1022	1069	945
Utah	Treatment	775	680	680
	Control	857	715	714
	Total	1632	1395	1394

Table 5. Sample descriptive for treatment and control group by study

Study	Group	Male	SPED	ELL	Ethnicity/ Race	Foster/ Homelessness
California	Control	47%	1%	35%	85% Hispanic	13%
	Treatment	53%	1%	28%	79% Hispanic	15%
Utah	Control	52%	8%	3%	82% White	N/A
	Treatment	51%	5%	5%	77% White	N/A

Analytical Approach

Hierarchical linear regression models (HLMs) with time (level 1) nested within students (level 2) were employed to examine growth in Composite and subscale scores. All models contained a series of covariates including gender (“female”; 1=female, 0=male), ethnicity (“hispanic”; 1= Hispanic, 0=not Hispanic) or race (“White”; 1= White, 0=not White), ELL status (“ELL”; 1=ELL, 0=non-ELL), SPED status (“SPED”; 1=SPED, 0=non-SPED), an indicator of fostering/homelessness (“foshom”; 1= in foster care or homeless, 0=not in foster care or homeless), an indicator of time (“Time”; 1=BOY, 2=MOY), an indicator of whether the student was in the treatment or control group (“intervention”; 1=treatment, 0=control), and an interaction between time and group calculated as the product of time*group (“Tigr”). The covariates used depended on availability by the district. When possible, among Below and Well Below Benchmark students, we also controlled for Phonological Awareness Core Survey total scores (“PATotal”) for kindergarten students.

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 or subscale score is different for the treatment versus control groups. We also looked at growth in Composite scores separately based on students’ benchmark scores.

We additionally explored the main effects of time (BOY vs. MOY) to measure the correlational impact of PA Lessons with all available statistical controls. All analyses were conducted separately by grade using the statistical software package R 3.6.2.

California Study MOY-BOY Statistical Results

All Students

In California, we examined growth in Composite scores as well as growth in FSF and LNF scores from BOY to MOY; PSF and NWF were not included in this analysis because neither measure is administered at BOY. Because the scores analyzed were highly positively 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=1.14, $p<.001$, $f^2=.00$) and FSF scores (IRR=1.12, $p<.001$, $f^2=.00$) across all students, regardless of whether they received the PA Lessons (Table 6). There was no effect of treatment on LNF scores. The results are displayed in Table # and the complete output for each model can be found in [Appendix 1](#).

Table 6. California Results for All Students

Test	School Group	BOY	MOY	Statistically Different?
Composite Scores	Wonders + Variety	24.63	113.30	Yes, schools using PA Lessons gained more.
	Wonders + PA Lessons	22.67	118.87	
First Sound Fluency Scores	Wonders + Variety	10.23	28.88	Yes, schools using PA Lessons gained more.
	Wonders + PA Lessons	9.54	30.14	
Letter Naming Fluency Scores	Wonders + Variety	15.65	34.72	No, they are similar. Treatment group saw significant growth from BOY to MOY.
	Wonders + PA Lessons	27.14	44.77	

Student BOY Groups: At/Above Benchmark and Below/Well Below Benchmark

We looked separately at students who were 1) At or Above Benchmark at BOY based on the composite (did not receive PA Lessons in either group) or 2) Below or Well Below Benchmark at BOY based on their composite (received PA Lessons in the treatment group). We found that students who were At or Above Benchmark in the treatment group demonstrated significant increases in MOY Composite scores (IRR=1.15, $p<.001$, $f^2=.00$), and students Below or Well Below Benchmark in the treatment group demonstrated marginally significant growth in MOY Composite scores (IRR=1.07, $p=.05$, $f^2=.00$). The results are displayed in Table 7 and the complete output for each model can be found in [Appendix 1](#).

Table 7. Composite Score Results, Differences by BOY Benchmark Group

Benchmark Group at BOY	School Group	BOY	MOY	Statistically Different?
At/Above Benchmark at BOY (Note: At/Above students did not receive intervention lessons)	Wonders + Variety	50.75	166.50	Yes, 95 PA Lessons gained more
	Wonders + PA Lesson	46.25	173.99	
Below or Well Below Benchmark at BOY	Wonders + Variety	6.09	61.19	Yes, 95 PA Lessons gained more
	Wonders + PA Lessons	5.60	60.10	

Utah Study MOY-BOY Results

Because the distribution of the Composite, FSF, and LNF scores were positively skewed within the kindergarten sample, we elected to use a Poisson distribution to examine changes in scores over time.

There was not a significant effect of treatment on kindergarten composite scores, suggesting that students in the treatment and control groups demonstrated similar growth. We looked separately at growth in composite scores among students who were 1) below or well below benchmark at baseline or 2) at or above benchmark at baseline. Among students who were below or well below benchmark at BOY, students in the control group demonstrated more growth in composite scores than students in the treatment group (IRR=0.94, $p=.015$, $f^2=.00$). Among students who were at or above benchmark at BOY, students in the treatment group demonstrated more growth in composite scores than students in the control group (IRR=1.02, $p=.03$, $f^2=.00$). See Tables 8 and 9 for predicted mean scores for all group combinations.

There was a significant effect of treatment on LNF scores; students in the control group demonstrated more growth in LNF scores than students in the treatment group (IRR=0.96, $p=.004$, $f^2=.00$; see Figure 1c). The complete output for each model can be found in [Appendix 2](#).

Table 8. Utah Results for All Students

Test	Group	BOY	MOY	Statistically Different?
Composite Scores	Journeys + Variety	30.7	133	No, they are similar. The treatment group saw significant growth from BOY to MOY.
	Journeys + PA Lessons	32.45	137.94	
First Sound Fluency Scores	Journeys + Variety	14.85	35.45	No, they are similar. The treatment group saw significant growth from BOY to MOY.
	Journeys + PA Lessons	15.33	36.58	
Letter Naming Fluency Scores	Journeys + Variety	10.95	24.66	Yes, Variety gained more
	Journeys + PA Lessons	8.04	11.78	

Table 9. Utah Composite Score Results, Differences by BOY Benchmark Group

Benchmark Group at BOY	Group	BOY	MOY	Statistically Different?
At/Above Benchmark at BOY	Journeys + Variety	48.38	155.87	Yes, PA Lessons gained more
	Journeys + PA Lessons	48.86	161.26	
Below or Well Below Benchmark at BOY	Journeys + Variety	7.74	77.94	Yes, Variety gained more
	Journeys + PA Lessons	7.92	75.41	

Key Findings



Treatment vs. Comparison Schools

When comparing treatment to comparison schools, the California study showed a more positive and significantly stronger impact of the PA Lessons on student outcomes than the variety of interventions used in the comparison schools.

Two reasons come to mind as to why the Utah school has not yet seen the full impact of their PA Lesson implementation.

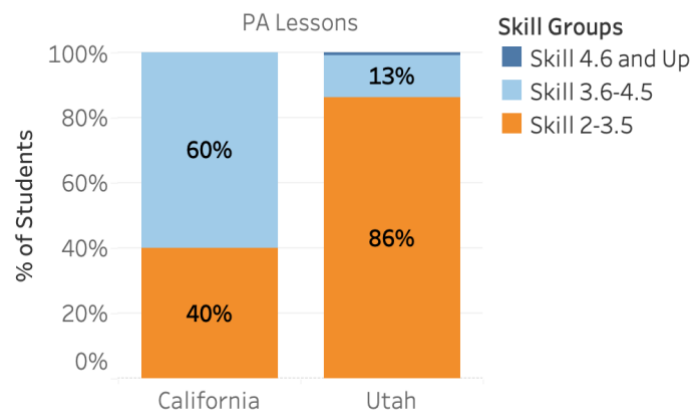
1. Most of students in Utah started right at the beginning of the PA Skill Continuum (86%), while students in California had 40% in early skills and 60% in Skills 3.6-4.5. It is more difficult to move students ahead when more of them are far behind.
2. More students in California than in Utah have covered all of the necessary lessons and have completed the PA Lesson skill continuum. Coaches visited both California and Utah in Winter 2022 to support and improve implementation for the spring semester.

Study	Gains on Composite	Gains on Subtests
California	+	+
Utah	→	→

 PA Lesson Group Higher Gains
  Similar Positive Gains

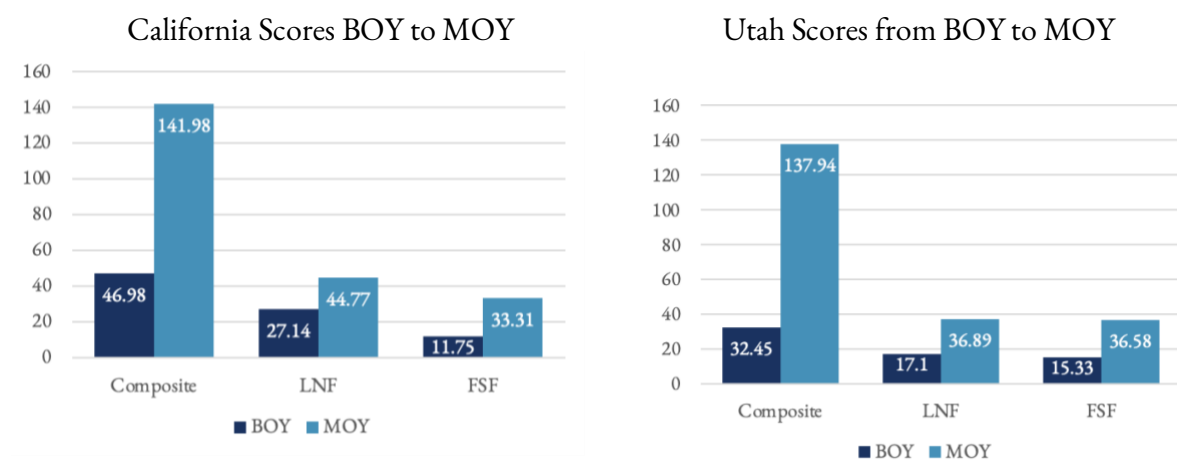
Study Site	Placed out by Cycle 5
California	50 (23%)
Utah	15 (4%)
Total	65 (10%)

Cycle 2 Results



Treatment Schools Only

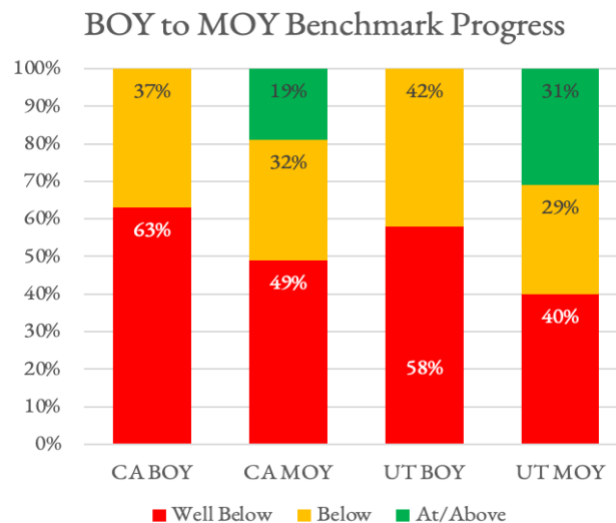
Students in PA Lessons schools made significant and positive gains from the beginning to the middle of the year in both studies.



Students using PA Lessons made progress and advanced benchmark categories in just a few months.

In California, 19% of Below and Well Below Benchmark students advanced to At/Above benchmark.

In Utah, 31% of Below and Well Below Benchmark students advanced to At/Above benchmark.



Conclusion and Next Steps

Conducting a large study in two districts allows for the examination of the PA Lessons with two student populations with different needs and backgrounds. Seeing positive effects, although quite small after only a few months of school, is very encouraging. Evidence of the program's impact was seen through the PA Lessons group outperforming the comparison groups on the Acadience Reading Composite scores and the First Sound Fluency scores in California. This positive impact was seen for all students and for students Below and Well Below Benchmark. Notably, California students who started the year At or Above Benchmark received benefits from the walk-to-intervention model, even though they did not receive the PA Lessons themselves. Perhaps what teachers learned from implementing the PA Lessons carried over into their core instruction. It's not unexpected that there was no difference in Letter Naming Fluency scores in California because the target of the PA Lesson intervention was phonological awareness and did not directly instruct letter names. It's likely that the variety of materials in the control group covered letter names more extensively.

The California district had a stronger start to the school year and more organized leadership compared to the Utah district, evidenced by its progress in accelerating students through the PA Skill Continuum faster than Utah. That advantage may partially explain why the California treatment schools outperformed the comparison schools consistently while the Utah treatment schools saw similar gains in composite scores as the comparison schools. A large number of treatment students in Utah did seem to gain a meaningful benefit from the PA lessons, with 31% of the Below and Well Below Benchmark students advancing to At/Above benchmark in just a few short months.

The major limitation of this study is that it only reviews a half year of data for a program that is intended to be used for a full year. The second challenge is that in Fall 2021, many students may have been absent from school due to COVID, without the options for remote schooling (attendance logs are not included in this analysis but will be included in the year-long study). Lastly, changing the model of intervention in a school from pull-out to walk-to-intervention takes many months to adopt and become routine. In addition, teaching PA explicitly may have been new for many teachers, so developing knowledge and facility with teaching these skills takes time. As with any large change, it takes time to get everything figured out. It is anticipated that the full-year study will be a more accurate reflection of the potential impact of the 95 Percent Group's Phonological Awareness Lessons.

The next steps for this study will be to conduct the Core Survey and Acadience end-of-year testing in April and May when the school districts complete their eleventh cycle of the PA Lessons.

Appendices

Appendix 1: California Results

- Composite score: (IRR=1.14, $p < .001$) - significant differences between treatment and control group
- FSF score: (IRR=1.12, $p < .001$) - significant differences between treatment and control group
- LNF score: (IRR=1.00, $p = .88$) - no significant differences between treatment and control group

For at or above students:

- Composite score: (IRR=1.15, $p < .001$) - significant differences between treatment and control group

For below or well below students:

- Composite score: (IRR=1.07, $p = .05$) - marginally significant differences between treatment and control group



Composite Score

<i>Predictors</i>	comp		
	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	6.64	5.81 – 7.59	<0.001
Time	4.04	3.96 – 4.11	<0.001
female	1.07	0.98 – 1.18	0.124
hispanic	0.91	0.81 – 1.03	0.133
ELL	0.82	0.74 – 0.91	<0.001
sp	0.50	0.34 – 0.73	<0.001
foshom	0.88	0.78 – 1.00	0.055
intervention	0.81	0.73 – 0.90	<0.001
Tigr	1.14	1.11 – 1.17	<0.001
Random Effects			
σ^2	0.02		
τ_{00} ID	0.49		
ICC	0.97		
N _{ID}	921		
Observations	1837		
Marginal R ² / Conditional R ²	0.524 / 0.984		

At or Above Benchmark Comparisons

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.

<i>Predictors</i>	comp		
	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	19.49	17.98 – 21.13	<0.001
Time	2.86	2.80 – 2.93	<0.001
female	0.98	0.93 – 1.04	0.498
hispanic	0.92	0.86 – 0.98	0.014
ELL	0.95	0.89 – 1.02	0.161
sp	0.89	0.67 – 1.17	0.393
foshom	1.00	0.93 – 1.09	0.933
intervention	0.79	0.73 – 0.86	<0.001
Tigr	1.15	1.11 – 1.18	<0.001
Random Effects			
σ^2	0.01		
$\tau_{00 ID}$	0.07		
ICC	0.88		
N_{ID}	404		
Observations	805		
Marginal R^2 / Conditional R^2	0.795 / 0.975		

Below Benchmark Comparisons

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.

<i>Predictors</i>	comp		
	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.70	0.56 – 0.89	0.003
Time	9.42	8.96 – 9.90	<0.001
female	1.21	1.06 – 1.39	0.006
hispanic	1.23	1.00 – 1.50	0.047
ELL	0.85	0.73 – 0.99	0.032
sp	0.55	0.32 – 0.94	0.028
foshom	0.88	0.73 – 1.06	0.185
PATotal	1.06	0.98 – 1.15	0.125
intervention	0.86	0.71 – 1.04	0.119
Tigr	1.07	1.00 – 1.14	0.054
Random Effects			
σ^2	0.03		
$\tau_{00 ID}$	0.48		
ICC	0.95		
N_{ID}	406		
Observations	810		
Marginal R^2 / Conditional R^2	0.725 / 0.986		

FSF

<i>Predictors</i>	fsf		
	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	4.35	3.75 – 5.05	<0.001
Time	2.52	2.45 – 2.60	<0.001
female	1.15	1.04 – 1.26	0.007
hispanic	0.95	0.84 – 1.09	0.469
ELL	0.80	0.71 – 0.89	<0.001
sp	0.47	0.31 – 0.71	<0.001
foshom	0.90	0.78 – 1.03	0.110
intervention	0.83	0.74 – 0.94	0.004
Tigr	1.12	1.07 – 1.17	<0.001
Random Effects			
σ^2	0.05		
$\tau_{00 \text{ ID}}$	0.54		
ICC	0.91		
N_{ID}	921		
Observations	1842		
Marginal R^2 / Conditional R^2	0.310 / 0.936		

LNF

<i>Predictors</i>	lnf		
	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	7.66	6.60 – 8.91	<0.001
Time	2.22	2.16 – 2.28	<0.001
female	0.97	0.88 – 1.07	0.542
hispanic	0.81	0.71 – 0.92	0.002
ELL	0.81	0.73 – 0.91	<0.001
sp	0.60	0.39 – 0.92	0.019
foshom	0.87	0.76 – 1.00	0.051
intervention	0.86	0.77 – 0.98	0.019
Tigr	1.00	0.96 – 1.04	0.882
Random Effects			
σ^2	0.05		
τ_{00} ID	0.58		
ICC	0.92		
N _{ID}	921		
Observations	1841		
Marginal R ² / Conditional R ²	0.231 / 0.936		



Appendix 2: Utah Results

- Composite score: (IRR=0.94, $p<.001$) - significant differences between treatment and control group

For at or above students:

- Composite score: (IRR=0.99, $p=.43$) - no significant differences between treatment and control group

For below or well below students:

- Composite score: (IRR=0.89, $p<.001$) - significant differences between treatment and control group
-

Composite Score

<i>Predictors</i>	comp score		
	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	4.70	4.30 – 5.15	<0.001
Time	4.56	4.49 – 4.63	<0.001
Gender	1.02	0.96 – 1.09	0.491
White	1.29	1.19 – 1.40	<0.001
ELL	0.69	0.59 – 0.82	<0.001
SPED	0.64	0.57 – 0.73	<0.001
intervention	1.16	1.07 – 1.24	<0.001
Tigr	0.94	0.92 – 0.96	<0.001
Random Effects			
σ^2	0.01		
τ_{00} STUDENTID	0.33		
ICC	0.96		
N STUDENTID	1393		
Observations	2786		
Marginal R ² / Conditional R ²	0.628 / 0.985		

At or Above Benchmark Comparisons

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.

<i>Predictors</i>	comp score		
	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	13.92	13.17 – 14.72	<0.001
Time	3.41	3.35 – 3.47	<0.001
Gender	0.99	0.96 – 1.02	0.356
White	1.02	0.98 – 1.06	0.379
ELL	0.93	0.82 – 1.05	0.232
SPED	0.88	0.82 – 0.95	0.001
intervention	1.06	1.00 – 1.11	0.039
Tigr	0.99	0.97 – 1.01	0.427
Random Effects			
σ^2	0.01		
τ_{00} STUDENTID	0.04		
ICC	0.83		
N STUDENTID	915		
Observations	1508		
Marginal R ² / Conditional R ²	0.873 / 0.978		

Below Benchmark Comparisons

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.

<i>Predictors</i>	comp score		
	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.80	0.70 – 0.92	0.001
Time	9.27	8.88 – 9.68	<0.001
Gender	1.03	0.95 – 1.12	0.478
White	1.24	1.12 – 1.37	<0.001
ELL	1.00	0.83 – 1.19	0.966
SPED	0.79	0.68 – 0.92	0.002
intervention	1.20	1.05 – 1.37	0.009
Tigr	0.89	0.83 – 0.94	<0.001
Random Effects			
σ^2	0.03		
τ_{00} STUDENTID	0.34		
ICC	0.91		
$N_{\text{STUDENTID}}$	800		
Observations	1278		
Marginal R^2 / Conditional R^2	0.759 / 0.979		

