



LXDRESEARCH
AT CHARLES RIVER MEDIA

Spalding Lookback 22-24 Efficacy Study Report

Examining the Impact of Spalding's Writing Road to Reading in Grades K-5 Across Multiple Years of Implementation

Conducted by Rachel L. Schechter, Ph.D., Rachel Gross, Ph.D., & Laura Janakiefski, Ph.D.

Abstract

This two-year study examined the impact of Spalding's *The Writing Road to Reading* literacy program on student achievement across grades K-5 in 14 Arizona schools. Using a quasi-experimental matched design, researchers compared 687 students from three Spalding schools with 1,228 students from eleven comparison schools using Foundations. Data from FastBridge earlyReading, CBMreading, and the Arizona Academic Standards Assessment (AASA) for reading were analyzed. After matching students and establishing baseline equivalence, students using Spalding consistently outperformed comparison students when followed longitudinally from 2022-2023 to 2023-2024. Kindergarteners followed to first grade showed greater growth than comparison students ($\eta^2 = .02$), and showed growth exceeding expected national growth benchmarks. Students in grades 2-4 were followed into grades 3-5 and demonstrated significantly higher AASA scores than matched peers ($\eta^2 = 0.05 - 0.12$). Spalding schools maintained consistent proficiency rates year-over-year, while comparison schools showed declines. Results indicate Spalding's *The Writing Road to Reading* effectively improves literacy outcomes across elementary grade levels.

73 TV Place, Needham MA 02494

www.lxdresearch.com

1 (617) 862-9212



Table of Contents

Introduction	1
Context of the Problem	1
Program Description	1
Study Context	3
Methods	3
Sample Description	3
District Overview	4
Implementation Description	4
Assessment Descriptions	5
Analysis Plan	5
Matching Procedure & Description of the Matched Samples	6
Planned Statistical Analyses	7
Study Results	8
Kindergarten to 1st Grade Findings	8
Proficiency Rates by Spring: K to 1st Grade	8
Comparisons Following Students from Year 1 to Year 2: K to 1st Grade	9
2nd to 5th Grade Findings	11
Proficiency Rates on the Arizona State Test: 3rd to 5th Grade	11
Comparisons Following Students from Grades 2-4 in Year 1 to Grades 3-5 in Year 2	13
Discussion	14
Summary of Key Findings	14
Early Literacy Development (K-1)	15
Upper Elementary Performance (Grades 2-5)	16
Implications for Practice	16
Limitations and Future Directions	17
Conclusion and Next Steps	17
References	19
Appendix	21



Introduction

Context of the Problem

The public education system in the United States faces a complex crisis where teacher shortages and retention challenges intersect with stalled academic recovery from the COVID-19 pandemic. While staffing challenges persist across multiple subject areas (McVey & Trinidad, 2019), reading instructors in particular face mounting pressure to improve literacy rates amid inadequate support. Many teachers have reported considering leaving the profession, citing issues with compensation, unreasonable expectations, and inability to protect their well-being (Bryant et al., 2023), factors that directly impact student achievement and school improvement efforts (Kini & Podolsky, 2016; Ronfeldt et al., 2013).

The teacher retention challenges are compounded by recent research that has revealed concerning patterns in student achievement that have lasted four years now after initial school closures from the pandemic. The data shows that students are either growing at slower rates than their historical peers or showing comparable growth but starting significantly behind, failing to make up lost ground from pandemic disruptions (Young & Young, 2024a; 2024b). This stagnant and, in some cases, declining recovery as of 2024 at the national level is evidenced by achievement data that largely mirrors 2022 and 2023 results, with only small increases in grade-level achievement in a few areas (Curriculum Associates, 2023; 2024; Lewis & Kuhfeld, 2024).

These challenges highlight the critical need for comprehensive approaches to literacy instruction and teacher support. While well-designed training and mentoring programs can improve teacher retention and efficacy (Sutcher et al., 2019; Geiger & Pivovarova, 2018), students simultaneously need well-designed literacy curricula with direct, explicit instruction in integrated spelling, writing, and reading skills. Many literacy programs fail to effectively combine these elements, and schools often lack the resources and teacher training necessary to deliver effective instruction that could help address pandemic-related learning gaps. This situation is particularly concerning as education leaders face the expiration of COVID-19 relief funds while still working to gain clear and accurate assessments of student achievement and implement effective recovery strategies.

Program Description

Spalding's *The Writing Road to Reading* is a literacy program designed to be used with its corresponding professional learning, with the aim of introducing students to the most

fundamental pieces of language. The program is designed to support teachers in delivering explicit, interactive, diagnostic, sequential, multi-modal, integrated instruction on spelling, writing, and reading. The program is informed by research in effective reading instruction, covering phonemic awareness, feature recognition, letter recognition, sound-symbol relationships, spatial placement, vocabulary, sentence structure, and text comprehension (e.g., Moats, 2020).

The program is often implemented at the whole school and whole class level, though individual teachers can be trained to use the program on their own in individual classrooms. The program is designed to empower teachers through the corresponding professional development to help them understand and apply the teaching philosophy known as The Spalding Method.

Spalding's *The Writing Road to Reading* is based on an instructional model that follows a cycle of modeling, coaching, and scaffolding to help students develop integrated skills through observation and guided practice (Collins, 1989; Collins et al., 1991). Research shows that modeling helps students understand cognitive tasks by enabling them to observe and build mental pictures of processes (Harbour et al., 2015), while scaffolding and targeted feedback help students improve their performance through specific, corrective guidance (Fyfe et al., 2022).

The Writing Road to Reading's instructional focus centers on three key research-backed approaches. First, it uses a speech-to-print methodology where learning moves from spoken word to printed word, helping students connect familiar spoken language with visual word forms (Moats, 2020). Second, it emphasizes feature and letter recognition, teaching students to distinguish vertical, diagonal, and horizontal lines as well as curves—skills that research shows are fundamental to letter recognition and early reading (Adams, 2011; Reutzel et al., 2017). Third, *The Writing Road to Reading* develops orthographic knowledge through explicit teaching of spelling patterns and rules, which has been shown to improve both spelling and reading skills (Squires & Wolter, 2016).

The effectiveness of *The Writing Road to Reading* is further supported by its emphasis on spaced practice and self-regulated learning. Research demonstrates that distributing practice across time rather than concentrating sessions leads to improved long-term retention (Carpenter et al., 2022; Latimier et al., 2021). Additionally, the program incorporates opportunities for student reflection and application, which helps develop self-regulated learning skills, which are linked to stronger decoding skills, reading comprehension, phonological awareness, and vocabulary



knowledge (Skibbe et al., 2019). The research-based foundation of *The Writing Road to Reading* has led to documented evidence of positive impacts on student learning, with studies showing schools implementing Spalding often outperforming local, state, and district averages in language arts, particularly in improving reading and spelling outcomes for students with dyslexia and other learning challenges. The current study aims to build on this previous work to understand the impact of Spalding's *The Writing Road to Reading* when implemented in multiple schools across multiple years.

Study Context

The present study was conducted by a team of independent researchers, [LXD Research](#), to examine the impact of Spalding's *The Writing Road to Reading* on student literacy achievement. Specifically, this study examines how students' growth is impacted by the implementation of Spalding's *The Writing Road to Reading* program within multiple schools, as well as how students using *The Writing Road to Reading* compare their literacy performance to students using comparison programs.

For this study, students were followed from year 1 of observation into year 2, and their year 2 scores were compared across treatment and comparison schools. Tracking students longitudinally from one year into the next and comparing their scores across conditions provides several key advantages for program evaluation. This approach accounts for the adaptation period educators and students need when adjusting to new learning methods, as initial performance may reflect adjustment difficulties rather than program effectiveness. Importantly, it also captures how teacher implementation proficiency may improve over time, as educators become more effective with a program as they gain experience. Longitudinal tracking generates more comprehensive evaluation data by revealing developmental trajectories that might be missed in single-year snapshots. Finally, this method captures cumulative and delayed effects of educational interventions, which may show stronger results after extended exposure as benefits compound or emerge only after sufficient implementation time.

Methods

Sample Description

1915 students across grades K to 4 were matched and longitudinally tracked across 2 years in grades 1 to 5. A total of 9647 student records were available in grades K-5 across the 14 schools that participated in the current study. 2133 student records came from one of 3 schools that

used Spalding for their literacy block, and 7514 student records came from one of 11 comparison schools that used comparison programs or materials. In the full sample of 2,725 students with data for Fall year 1 and Spring year 2, 2,072 (76.1%) were in the Comparison group, while 653 (23.9%) were in the Treatment group.

The total sample size (N = 9647) reflects the sum of the student counts used in the separated grade-level cohort analyses for consistency in reporting with how the analyses were completed. Because the analyses were split by grade level, some students could appear in both the 2022-23 and 2023-24 counts for the total sample size; for example, a student who was a kindergarten student in 2022-23 and a first grader in 2023-24 would be reflected in both counts for the total sample size. As such, the total sample size reflects the total count of student records across the unmatched sample analyses described below. However, for the matched sample and corresponding analyses described below, there were a total of 1915 students who were only counted once, and followed across years into the next grade.

Table 1. Sample Sizes by Grade, Year, and Condition for the Unmatched Sample

Grade	Condition	Sample Size					
		K	1st	3rd	4th	5th	Total
2022-2023	Treatment	175	221	246	194	226	1062
	Comparison	736	839	708	743	754	3780
2023-2024	Treatment	205	203	200	255	208	1071
	Comparison	746	819	702	737	730	3734
Total		1862	2082	1856	1929	1918	9647

District Overview

The participating district was a large district in Arizona that consists of 15 elementary schools, with a student body of over 10,000 students. There are over 78% Hispanic/Latino students, and over 78% of students in the district are eligible to participate in free and reduced price meals. The average student-teacher ratio is 18:1.

Implementation Description

Spalding's *The Writing Road to Reading* was implemented in 3 schools in the current study (i.e., 'Spalding schools'). There were an additional 11 schools serving as comparison schools. The Spalding schools had their teachers complete the Spalding training, and had between 2-5 trained Spalding coaches onsite to support implementation. Comparison schools used Foundations as their core literacy program during the 2 years observed in the current study. Spalding's *The Writing Road to Reading* was used as a whole-class program for literacy, implemented during the literacy block. Schools were followed for 2 years, 2022-2023 and 2023-2024, to assess year-over-year effects of continued implementation.

One of the Spalding schools has been using *The Writing Road to Reading* for over 40 years. The other 2 Spalding schools started using *The Writing Road to Reading* 2-3 years ago. All 3 Spalding schools had onsite Spalding coaches to support implementation throughout the year. Given that the Spalding schools had been using the program prior to the beginning of observation for the current study, matching procedures for the Spalding and comparison schools were followed in order to establish baseline equivalence for a controlled comparison between the two groups.

Assessment Descriptions

The participating school district uses FastBridge assessments for progress monitoring to understand student progress and performance across the year. The schools in this study used the [earlyReading](#) subtest for Kindergarten and 1st grade, and [CBMreading](#) for 2nd to 5th grade. FastBridge earlyReading and CBMreading measure students' proficiency in essential early reading skills such as concepts of print, phonemic awareness, phonics, and fluency. FastBridge earlyReading is the K-1 assessment companion to CBMreading. Both assessments are curriculum-based measures that are teacher-administered at 3 different time points during the school year: Fall, Winter, and Spring. earlyReading composite scores and CBMreading words read correct per minute (WRCPM) were used in the analyses for comparison across conditions as well as reference to the FastBridge norming studies that determined [percentile cutpoints](#).

Students in grades 3-8 also completed Arizona's Academic Standards Assessment (AASA), which is the statewide achievement test administered in the spring. The ELA portion of the assessment addresses key components of reading and writing according to the ELA Standards adopted by the Arizona State Board of Education in 2016. Items on the assessment range from basic tasks such as correcting grammatical and spelling errors and defining terms to more complex, higher-

order tasks such as assessing logical arguments and making inferences based on texts while citing evidence to support claims. For more information about how the items for each grade level were determined, see these ELA Item Specification Overviews: [Grade 3](#), [Grade 4](#), [Grade 5](#). Students receive composite scores that are categorized into four performance levels based on [cut scores](#): Minimally Proficient (Level 1), Partially Proficient (Level 2), Proficient (Level 3), and Highly Proficient (Level 4). Each grade has a detailed set of Performance Level Descriptors (see [Grade 3](#), [Grade 4](#), [Grade 5](#)). Passing Proficiency levels include Proficient or Highly Proficient scores, as students who score in these ranges are likely to be ready for the next grade.

Analysis Plan

First, baseline equivalence was assessed at the first measurement timepoint, Fall 2022. Baseline equivalence was not met across the Spalding and comparison schools for any grade level, such that the Spalding schools were outperforming the comparison schools. To establish baseline equivalence between groups, the MatchIt package (Ho, Imai, King, & Stuart, 2011; Ho, Imai, King, Stuart, & Greifer, 2024) in R was used on the composite scores for the earlyReading assessment for K-1 and CBMReading for grades 2-4 to establish grade-level groups with similar baseline scores.

Matching Procedure & Description of the Matched Samples

The matching procedure for Kindergarten involved students from the 2022–23 cohort with data for the Fall 2022 earlyReading English Composite Score. Matching took place via the MatchIt package, where propensity scores were estimated using a logistic regression with the Fall 2022 earlyReading English Composite Score. The matching resulted in 146 pairs (146 treatment and 146 comparison cases) drawn from an initial pool of 163 Spalding students and 644 comparison students. After matching, both groups exhibited identical means on the outcome and quantile category, with a standardized mean difference (SMD) of 0, meeting the ESSA baseline equivalence criterion of $SMD \leq .25$.

The matching procedure for grades 2, 3, and 4 involved students from the 2022–23 cohort with Fall 2022 CBMR data for English Median WRC (words read correct per minute). Matching took place with the MatchIt package, where propensity scores were estimated via logistic regression with the CBMreading English Median WRC score. For grade 2, this yielded 160 treatment students matched to 320 comparison students with an SMD of .067, meeting the ESSA baseline equivalence criterion of an $SMD \leq .25$. For grade 3, the procedure matched 209 treatment

students with 418 comparison students, with an SMD of .124, meeting the ESSA baseline equivalence criterion. For grade 4, the matching procedure paired 172 treatment students with 344 matched comparison students, with an SMD of .058 meeting the ESSA baseline equivalence criterion.

Table 2. Demographic Information for the Matched Samples by Grade and Condition

Grade	Condition	Sample Size	Gender (% male)	Hispanic or Latino	Black or African American	Special Ed Status
K to 1	Treatment	146	55%	73%	6%	5%
	Comparison	146	55%	72%	12%	12%
2 to 3	Treatment	160	52%	75%	6%	6%
	Comparison	320	54%	78%	11%	10%
3 to 4	Treatment	209	50%	76%	6%	8%
	Comparison	418	50%	83%	7%	7%
4 to 5	Treatment	172	58%	76%	6%	8%
	Comparison	344	50%	82%	8%	5%
Total	Treatment	687	53%	75%	6%	7%
	Comparison	1228	52%	80%	9%	8%

Table 3. Descriptive Statistics for the Matched Samples by Grade and Condition

Grade	Condition	BOY Year 1 Measure	BOY Year 1 Mean	SD	EOY Year 2 Measure	EOY Year 2 Mean	SD
K to 1	Treatment	earlyReading	30.4	3.5	earlyReading	72.7	26.9
	Comparison		30.4	3.5		68.1	22.4
1 to 2	Treatment	Not trackable with available data					
	Comparison						
2 to 3	Treatment	CBMR	44.8	31.2	AASA	2499	31.2
	Comparison		43.0	29.6		2489	28.6
3 to 4	Treatment	CBMR	78.6	38.3	AASA	2525	35.3
	Comparison		74.7	36.0		2507	29.4
4 to 5	Treatment	CBMR	108.1	37.4	AASA	2535	32.8
	Comparison		105.4	37.2		2521	30.2

Statistical Analyses Plan

To understand how Spalding and comparison schools were performing in the unmatched sample, Chi-squared tests of independence were used to compare proficiency levels across schools and the two years observed to determine if the proportion of proficient students differed significantly. Follow-up chi-squared tests of independence were planned to compare Spalding students in the first year of observation versus the second year of observation (i.e., kindergarteners in 22-23 vs. kindergarteners in 23-24).

After establishing baseline equivalence in the matched sample, students' growth and performance across years was compared. Students in Spalding and comparison schools were followed from year 1 of observation to year 2 (e.g., matched in kindergarten and followed to first grade) and their performance was compared at the end of the second year of observation (i.e., Spring 2024). ANCOVA was used to compare year 2 performance across groups, controlling for baseline reading ability and demographics. Within the Spalding schools, K-1 students were also tracked to understand their year-over-year growth, which was descriptively compared to growth benchmarks (i.e., expected growth for students at the 15th percentile and the 40th percentile) according to the FastBridge earlyReading assessment. 2nd to 4th graders followed into 3rd to 5th grade did not have available data to track growth, as their baseline score measure (i.e., CBMreading) differed from their performance measure (i.e., AASA state test).

earlyReading composite scores were used for matching students and tracking students from Kindergarten through 1st grade. CBMreading Words Read Correct Per Minute (WRCPM) was used to match students in 2nd, 3rd, and 4th grade. Students were then followed to their AASA state test performance in the following year (e.g., 3rd, 4th, and 5th grade, respectively), which was compared across Spalding and comparison schools. There was no trackable data available for 1st graders moving into 2nd grade, because the progress monitoring measure changed between the two grades meaning there weren't assessments administered that could track growth from one year to the next.

Study Results

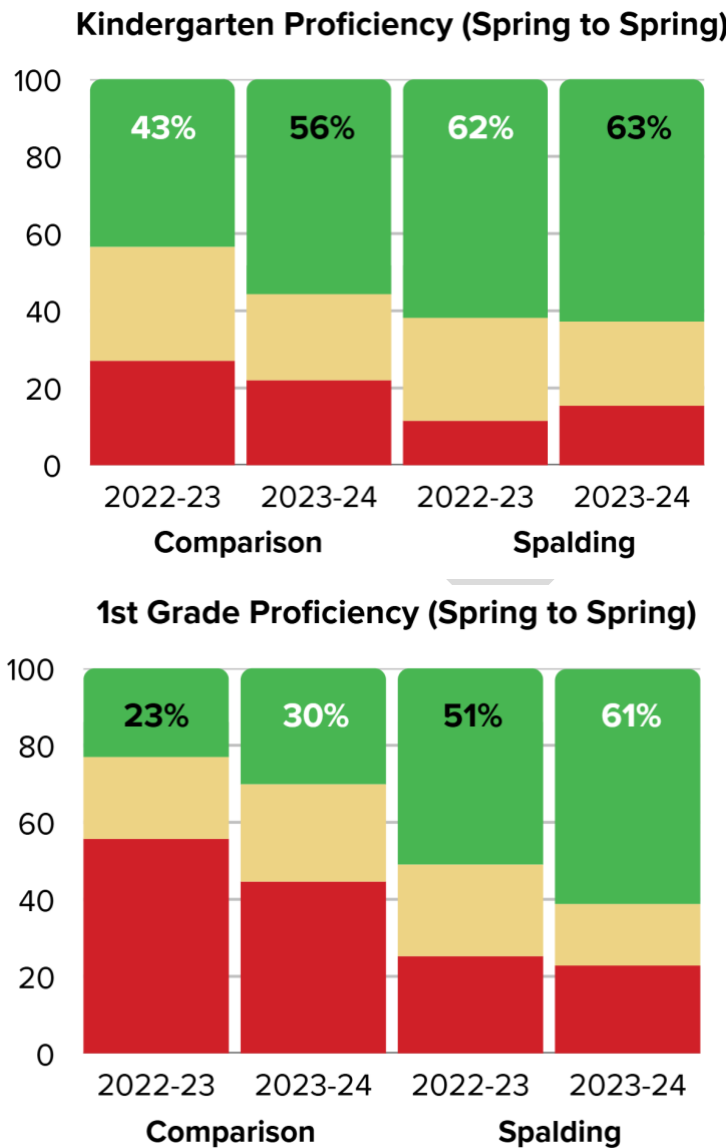
Kindergarten to 1st Grade Findings

Proficiency Rates by Spring: K to 1st Grade

The proportion of Kindergarten and 1st grade students meeting proficiency levels on earlyReading by Spring was compared across Spalding and comparison schools over 2 years in the unmatched sample using Chi-squared tests of independence. There was a significant difference in the relative distributions for both Kindergarten ($N = 1653$; $\chi^2(6) = 45.61$, $p < .001$), and First grade comparisons ($N = 1847$; $\chi^2(6) = 152.37$, $p < .001$). The distribution of students across the 3 proficiency levels differed significantly between comparison schools and Spalding schools across the two years examined.

Comparing grade-level cohorts within Spalding schools across the first year of observation versus the second year of observation (i.e., Kindergarteners in 22-23 vs. Kindergarteners in 23-24), Chi-squared tests of independence were used to determine whether the proportion of students meeting proficiency status (i.e., surpassing the 40th percentile benchmark, as defined by FastBridge) differed across years. Schools using Spalding's *The Writing Road to Reading* showed a reduction in the proportion of students performing below proficiency levels (e.g., High Risk and Some Risk) and an increase in students meeting proficiency standards (e.g., Low Risk, above the 40th percentile) from the first year observed to the second year observed. The proportion of students meeting proficiency across years was roughly equivalent for kindergarteners, $\chi^2(1) = 0.05$, $p = .82$, ϕ effect size = 0.01, but there was a significant shift in the proportion of proficient 1st graders, $\chi^2(1) = 3.91$, $p < .05$, ϕ effect size = 0.1. In addition, from spring 2023 to spring 2024, Spalding schools kept a larger proportion of their students proficient. Following numbers in white from the top graph to the bottom graph, Spalding schools maintain a consistent rate of proficiency, going from 62% proficient in 22-23 to 61% proficient in 23-24, whereas comparison schools drop from 43% proficient in 22-23 to 30% proficient in 23-24. See Figure 1.

Figure 1. Proficiency Levels for K-1 Spalding and Comparison school students

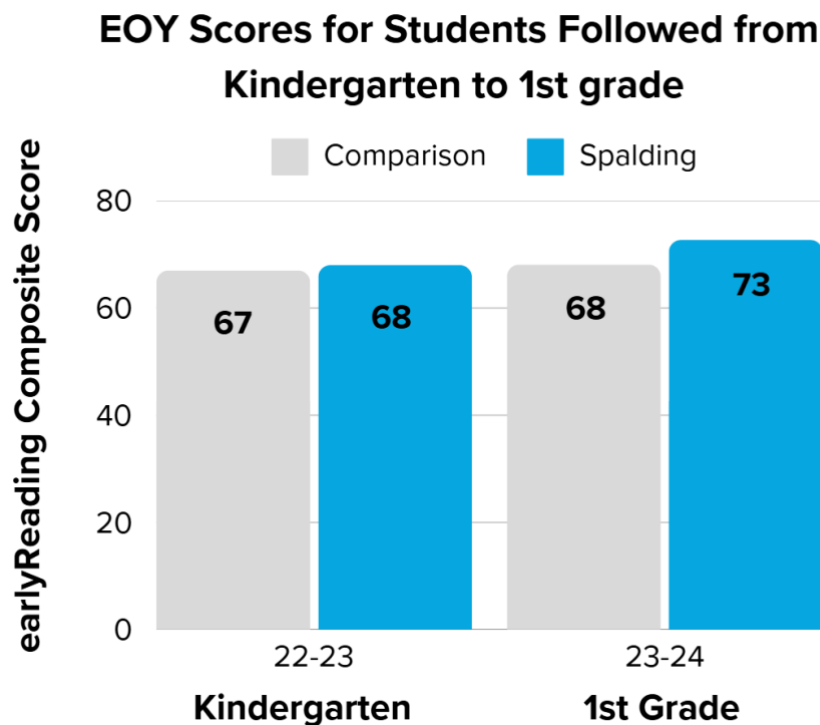


Note. Green indicates 'Proficient' (> 40th percentile on earlyReading), yellow indicates 'Some Risk' (between 15th percentile and 40th percentile), and red indicates 'High Risk' (below 15th percentile).

Comparisons Following Students from Year 1 to Year 2: K to 1st Grade

After matching students on baseline scores and following them from Kindergarten to 1st grade on the FastBridge earlyReading assessment (N = 292), ANCOVA was used to compare across Spalding and comparison schools, while controlling for baseline scores, gender, race, and special education status. Students using Spalding’s *The Writing Road to Reading* had higher Spring 2024 earlyReading scores than comparison students, $F(1, 238) = 4.63, p < .05, \eta^2 = .02$, small to medium effect. In other words, there was a small to medium effect of the program, such that students using Spalding’s *The Writing Road to Reading* who were determined to start in the same place as comparison students in kindergarten ended up showing higher scores at the end of 1st grade. See Figure 2.

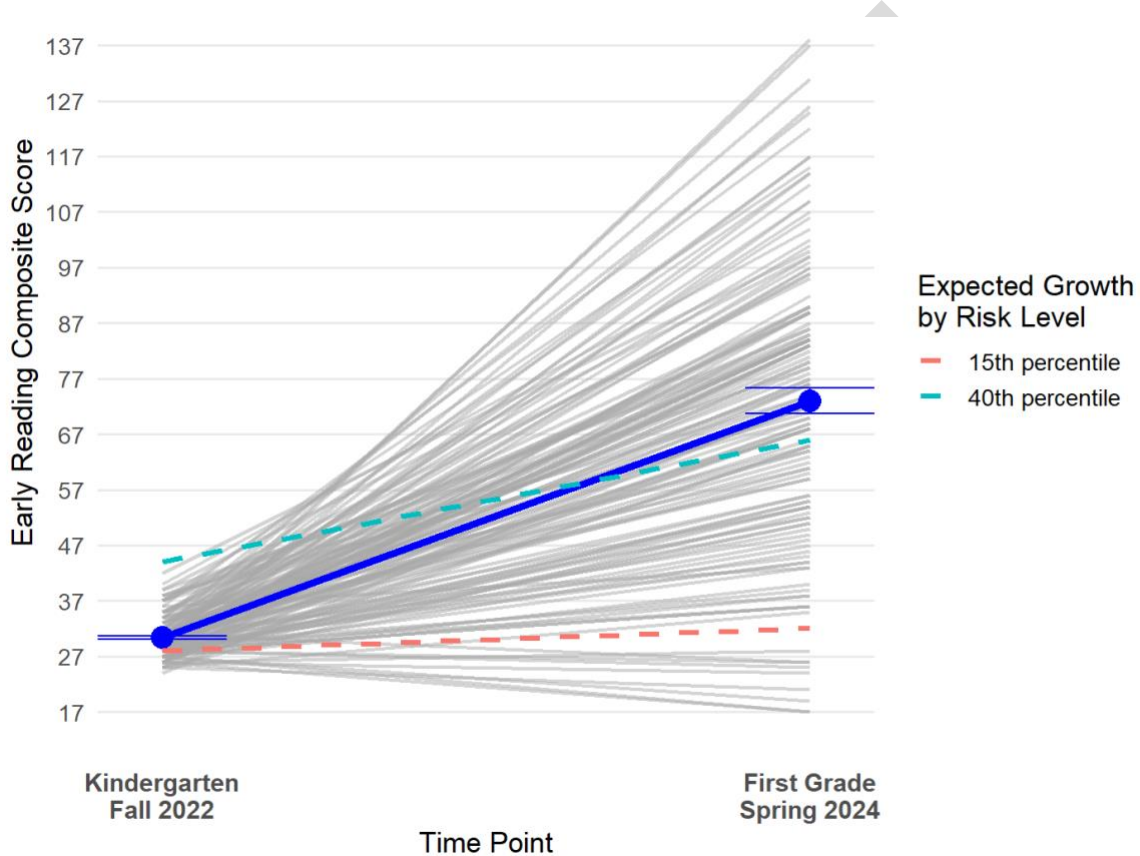
Figure 2. Spring earlyReading Scores for Students Followed from Kindergarten to 1st Grade



Focusing in on the Spalding schools only, ANCOVA was used to capture growth from Fall 22-23 to Spring 23-24, while controlling for baseline scores, gender, race, and special education status. There was a significant effect of time, indicating a significant score increase from the start of the first year to the end of the second year, $F(1, 130) = 355.8, p < .001, \eta^2 = .73$, large effect. As Figure 3 shows, Spalding students progressed significantly, outpacing the expected growth for

students at the 40th percentile, a key benchmark defined by the FastBridge assessment indicating that a student meets proficiency standards. In the matched sample, Spalding students started around the 15th percentile, yet outpaced the expected growth for students at the 40th percentile.

Figure 3. Spalding Student Growth Compared to National Norms of Expected Growth



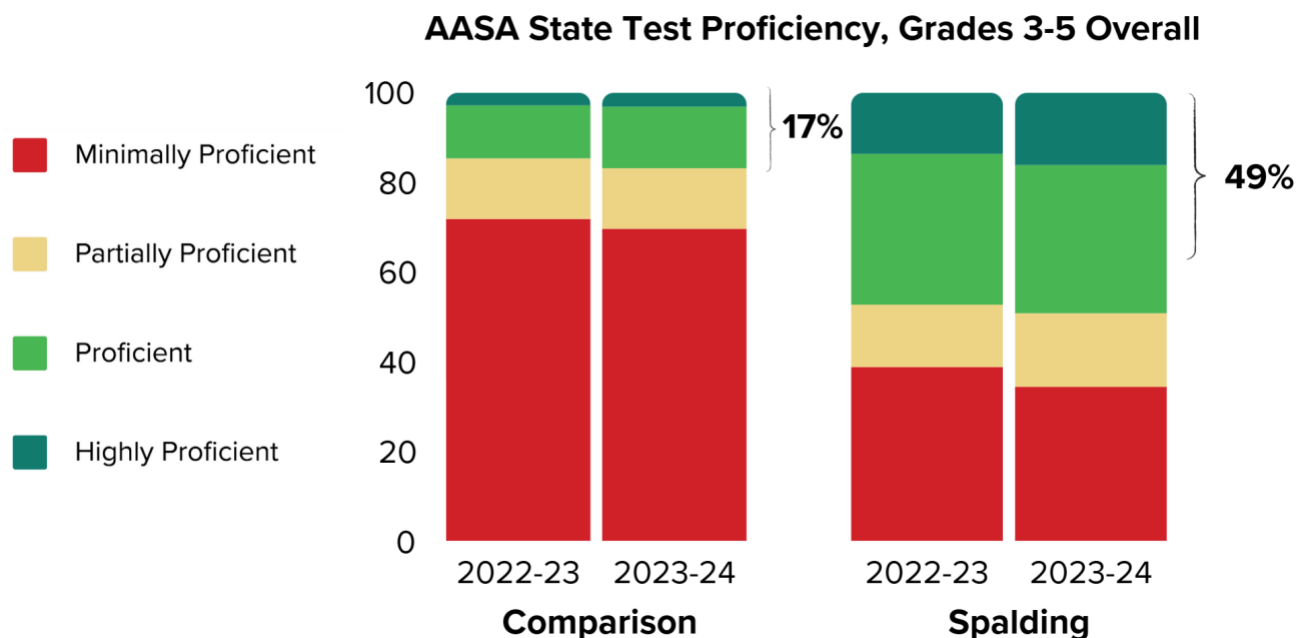
Note: The dashed pink line indicates the expected growth for students considered 'High Risk' by the FastBridge cut scores, scoring below the 15th percentile; the dashed blue line indicates the expected growth for students considered 'At Risk' by scoring below the 40th percentile benchmark.

2nd to 5th Grade Findings

Proficiency Rates on the Arizona State Test: 3rd to 5th Grade

The proportion of 3rd through 5th grade students in the unmatched sample meeting proficiency levels on the Arizona state test (AASA) was compared across Spalding and comparison schools over 2 years using Chi-squared tests of independence. There was a significant difference in the distribution of proficiency rates across the two groups and school years observed ($N = 5703$; $\chi^2(9) = 695.7$, $p < .001$). In addition, there were fewer students performing below proficiency levels and more students meeting proficiency benchmarks in Spalding schools than comparison schools during the 2022-2023 school year, Spalding 22-23 ($n = 666$) vs. Comparison 22-23 ($n = 2205$): $\chi^2(3) = 343.01$, $p < .001$, and this was also seen for the 2023-2024 school year comparison, Spalding 23-24 ($n = 663$) vs. Comparison 23-24 ($n = 2169$): $\chi^2(3) = 347.03$, $p < .001$. There was not a significant difference in the distribution of students across proficiency levels for the Spalding 22-23 and Spalding 23-24 comparison, $\chi^2(3) = 4.46$, $p = .21$, Cramer's V effect size = 0.03.

Figure 4. AASA State Test Proficiency Rates for Grades 3-5 Combined



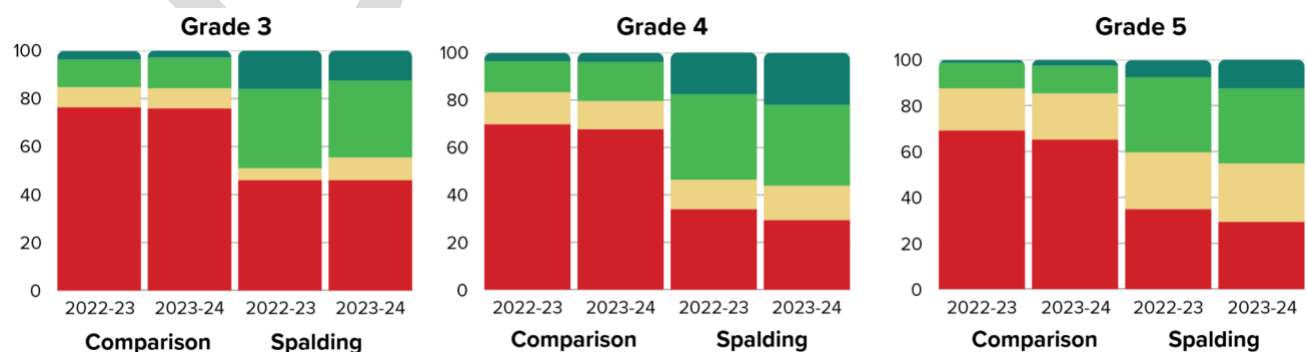
This pattern also held for each of the grades 3-5 separately. Across 2022-23 and 2023-24, there were more students meeting proficiency benchmarks in Spalding schools than comparison schools.

For Grade 3, the overall distribution across conditions and years was significantly different, $\chi^2(9) = 204.99$, $p < .001$, Cramer’s V effect size = 0.19. There was a significant difference between Spalding and comparison schools for 2022-23, $\chi^2(3) = 117.05$, $p < .001$, Cramer’s V effect size = 0.35. There was also a significant difference between Spalding and comparison schools for 2023-24, $\chi^2(3) = 82.93$, $p < .001$, Cramer’s V effect size = 0.30. There was not a significant difference in the distribution of students across proficiency levels for the Spalding 22-23 and Spalding 23-24 comparison, $\chi^2(3) = 3.81$, $p = .28$, Cramer’s V effect size = 0.04.

For Grade 4, the overall distribution across conditions and years was significantly different, $\chi^2(9) = 282.24$, $p < .001$, Cramer’s V effect size = 0.22. There was a significant difference between Spalding and comparison schools for 2022-23, $\chi^2(3) = 124.84$, $p < .001$, Cramer’s V effect size = 0.36. There was also a significant difference between Spalding and comparison schools for 2023-24, $\chi^2(3) = 149.16$, $p < .001$, Cramer’s V effect size = 0.38. There was not a significant difference in the distribution of students across proficiency levels for the Spalding 22-23 and Spalding 23-24 comparison, $\chi^2(3) = 2.32$, $p = .51$, Cramer’s V effect size = 0.00.

For Grade 5, the overall distribution across conditions and years was significantly different, $\chi^2(9) = 234.34$, $p < .001$, Cramer’s V effect size = 0.20. There was a significant difference between Spalding and comparison schools for 2022-23, $\chi^2(3) = 111.55$, $p < .001$, Cramer’s V effect size = 0.33. There was also a significant difference between Spalding and comparison schools for 2023-24, $\chi^2(3) = 113.07$, $p < .001$, Cramer’s V effect size = 0.34. There was not a significant difference in the distribution of students across proficiency levels for the Spalding 22-23 and Spalding 23-24 comparison, $\chi^2(3) = 3.79$, $p = .28$, Cramer’s V effect size = 0.04.

Figure 5. AASA State Test Proficiency Rates for Grades 3-5, Split by Grade

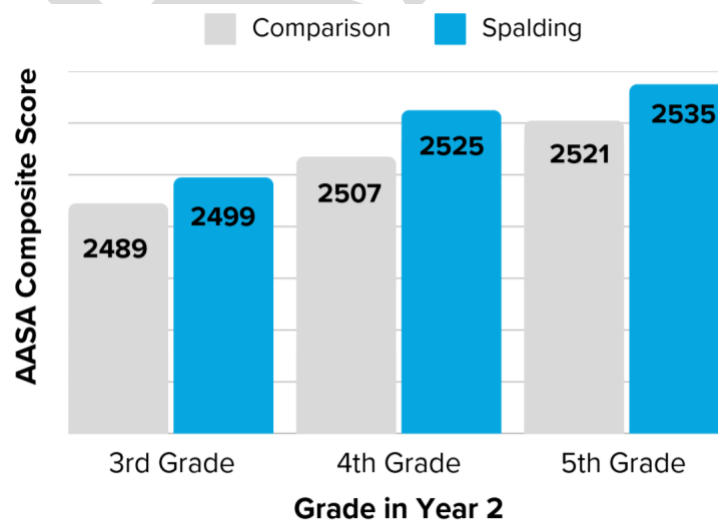


Comparisons Following Students from Grades 2-4 in Year 1 to Grades 3-5 in Year 2

After matching students in Spalding and comparison schools on their 2022-23 Fall scores for each grade level, their Arizona Academic Standards Assessment (AASA) scores were compared at the end of Year 2 of observation (e.g., 2nd graders were matched in Year 1, and then later their 3rd grade state test scores were compared at the end of Year 2, 2023-24).

Each grade level of Spalding students outperformed comparison students, controlling for baseline scores and demographic covariates, including gender, race, and special education status. For students matched in 2nd grade and compared on their 3rd grade state test scores (n = 480), Spalding students performed significantly better than comparison students on the state test, $F(1, 357) = 16.79, p < .001, \eta^2 = .05$, small to medium effect. For students matched in 3rd grade and compared on their 4th grade state test scores (n = 627), Spalding students performed significantly better than comparison students on the state test, $F(1, 476) = 61.99, p < .001, \eta^2 = .12$, medium to large effect. For students matched in 4th grade and compared on their 5th grade state test scores (n = 516), Spalding students performed significantly better than comparison students on the state test, $F(1, 393) = 33.64, p < .001, \eta^2 = .08$, medium to large effect. In other words, across each grade level, Spalding students who started at the same place as comparison students in the first year of observation showed higher scores at the end of the second year of observation. See Figure 6.

Figure 6. AASA State Test Scores for 3rd to 5th Grade Students in Year 2



Discussion

Summary of Key Findings

The results of this two-year study provide compelling evidence for the effectiveness of Spalding’s *The Writing Road to Reading* program in improving early literacy outcomes across elementary grade levels. The data demonstrate consistent advantages for students in Spalding schools compared to those in comparison schools, with effects observable from kindergarten through fifth grade. Overall, students in Spalding schools showed improved performance compared to schools using comparison programs (See Table 4 for a summary). Spalding schools also showed compounding growth across years of implementation, with higher scores and increased proficiency levels in the second year of observed implementation compared to the first.

Looking at the proficiency rate findings, there is a clear ‘flipping of the triangle.’ Unlike the comparison schools who had most of their students below proficiency, with only a few above proficiency, the Spalding schools had most of their students above proficiency levels, with fewer students falling below proficiency levels.

Table 4. Summary of Study Findings Split by Grade

Grade	Significant Difference vs. Comparison Schools	Outcome Measure
K to 1st	✓	FastBridge earlyReading
1st to 2nd	Not trackable with available data	
2nd to 3rd	✓	AASA
3rd to 4th	✓	AASA
4th to 5th	✓	AASA

Early Literacy Development (K-1)

The significant differences in proficiency distributions between Spalding and comparison schools in both kindergarten and first grade highlight the program's impact during the critical early literacy acquisition phase. Particularly noteworthy is the year-over-year improvement within Spalding schools themselves, with a significant increase in first-grade proficiency rates from the first to second year of implementation. This suggests that program effectiveness may strengthen as teachers gain experience with the methodology and as school-wide implementation becomes more established.

The longitudinal analysis following matched kindergarteners into first grade provides even stronger evidence for Spalding's effectiveness. After controlling for baseline scores and demographic factors, Spalding students demonstrated significantly higher earlyReading scores than their comparison school counterparts. This controlled comparison strengthens the case that observed differences can be attributed to the instructional approach rather than pre-existing student characteristics.

Perhaps most impressive is the growth trajectory observed in Spalding students. Starting around the 15th percentile nationally, these students not only made substantial gains but exceeded the expected growth patterns for students at the 40th percentile benchmark. This accelerated progress suggests that the Spalding method may be particularly effective at helping students overcome initial reading difficulties and build momentum in their literacy development.

Upper Elementary Performance (Grades 2-5)

The advantages of the Spalding method appear to extend well beyond early literacy foundations. The significant differences in state test proficiency rates between Spalding and comparison schools in both study years demonstrate that these early literacy gains translate into broader academic achievement as measured by standardized assessments.

The matched-cohort analyses following students from grades 2-4 into grades 3-5 are especially telling. With effect sizes ranging from moderate ($\eta^2 = .05$) to large ($\eta^2 = .12$), Spalding students consistently outperformed their matched peers from comparison schools across all three grade-level transitions. The largest effect was observed in the transition from third to fourth grade ($\eta^2 = .12$), suggesting that the Spalding method may have particular strengths in supporting the critical shift from "learning to read" to "reading to learn" that typically occurs around this age.

Implications for Practice

These findings have several important implications for educational practice:

1. **Early Implementation Benefits:** Schools adopting Spalding's *The Writing Road to Reading* may see benefits as early as the first year of implementation, with increasing returns in subsequent years as the program becomes more established.
2. **Closing Achievement Gaps:** The accelerated growth trajectory observed in Spalding students who began below proficiency benchmarks suggests this approach may be particularly valuable for supporting struggling readers and potentially narrowing achievement gaps.
3. **Sustained Impact:** The consistent advantages seen across grade levels indicate that the Spalding method builds foundational skills that continue to benefit students throughout their elementary education, rather than producing only short-term gains.
4. **Comprehensive Literacy Approach:** The positive impact on standardized test performance suggests that Spalding's *The Writing Road to Reading* develops not just discrete reading skills but broader literacy competencies that transfer to general academic achievement.

Limitations and Future Directions

Several important limitations should be considered when interpreting the findings of this study. First, changes in assessment protocols between grade 1 and grade 2 prevented continuous tracking of student progress across these critical early years. Second, baseline differences between groups presented methodological challenges. The Spalding group demonstrated significantly higher initial performance compared to the comparison group, necessitating the creation of a matched sample from a subset of students to establish baseline equivalence. While this matching process improved comparability, it reduced the effective sample size and potentially limits generalizability. Third, the retrospective design of this study meant that the research team could not control for various factors in how the program was implemented across different schools, introducing potential variability in program fidelity and implementation quality. Additionally, some Spalding schools had implemented the program for an extended period prior to the study. Despite the application of careful statistical controls in the quasi-experimental study, these factors preclude definitive causal attributions regarding program effectiveness.

Additionally, the study does not provide insight into which specific components of the Spalding method were most influential in driving student outcomes.

Future research should explore implementation factors that may moderate program effectiveness, examine longer-term impacts beyond elementary school, and investigate whether the Spalding method has differential effects for various student subgroups. A more detailed analysis of instructional practices within Spalding and comparison classrooms would also help identify the specific pedagogical mechanisms responsible for the observed advantages.

Conclusion and Next Steps

Overall, students in Spalding schools showed improved performance compared to schools using comparison programs. Spalding schools also showed compounding growth across years of implementation, with higher scores and increased proficiency levels in the second year of observed implementation compared to the first. Recommended next steps include following students through an additional year within Spalding schools to measure effects of continued implementation on year-over-year performance, and to capture educator voices in terms of implementation strengths and feedback.

The consistent pattern of results across multiple grade levels, measurement approaches, and cohorts provides strong evidence for the effectiveness of the Spalding method in promoting literacy development and academic achievement. The method appears to not only accelerate early reading acquisition but also build a foundation for sustained academic success throughout elementary school. These findings suggest that wider implementation of the Spalding approach represents an effective strategy for improving literacy outcomes in elementary education.

References

- Adams, M. J. (2011). The relation between alphabetic basics, word recognition, and reading. In S.J. Samuels & A.E. Farstrup (Eds.), *What Research Has to Say About Reading Instruction* (pp. 4-24). Newark, DE: International Reading Association.
- Bryant, J., Ram, S., Scott, D., & Williams, C. (2023). Education Practice K-12 teachers are quitting. What would make them stay?
<https://www.mckinsey.com/industries/education/our-insights/k-12-teachers-are-quitting-what-would-make-them-stay>
- Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, 1(9), 496–511.
<https://doi.org/10.1038/s44159-022-00089-1>
- Collins, A. (1989). Technical Report No. 474 Cognitive Apprenticeship And Instructional Technology. BBN Laboratories. <https://eric.ed.gov/?id=ED331465>
- Collins, A., Seely Brown, J., & Holum, A. (1991). Cognitive Apprenticeship. *American Educator*, Winter. https://www.aft.org/ae/winter1991/collins_brown_holum
- Curriculum Associates. (2023). State of student learning in 2023.
<https://cdn.bfldr.com/LS6J0F7/at/x8v8wp2c6j4s4wttsw2nwphb/ca-state-of-studentlearning-technical-report-2023.pdf>
- Curriculum Associates (2024). State of student learning in 2024.
<https://cdn.bfldr.com/LS6J0F7/at/wxj37b648k5bkwvf3vrrgj8/2024-State-of-Student-Learning-Research-Awareness-Technical-Report.pdf>
- Fyfe, E. R., Borriello, G. A., & Merrick, M. (2022). A developmental perspective on feedback: How corrective feedback influences children’s literacy, mathematics, and problem solving. *Educational Psychologist*, 58(3), 1–16.
<https://doi.org/10.1080/00461520.2022.2108426>
- Geiger, T., & Pivovarova, M. (2018). The effects of working conditions on teacher retention. *Teachers and Teaching*, 24(6), 604-625.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). *MatchIt: Nonparametric preprocessing for parametric causal inference*. *Journal of Statistical Software*, 42(8), 1–28.
- Ho, D. E., Imai, K., King, G., Stuart, E. A., & Greifer, N. (2024). *MatchIt: Nonparametric preprocessing for parametric causal inference* [Computer software].
<https://doi.org/10.32614/CRAN.package.MatchIt>
- Harbour, K. E., Evanovich, L. L., Sweigart, C. A., & Hughes, L. E. (2015). A Brief Review of Effective Teaching Practices That Maximize Student Engagement. *Preventing School*

- Failure: Alternative Education for Children and Youth*, 59(1), 5–13.
<https://doi.org/10.1080/1045988x.2014.919136>
- Kini, T., & Podolsky, A. (2016). Does Teaching Experience Increase Teacher Effectiveness? A Review of the Research. Learning Policy Institute.
- Latimier, A., Peyre, H., & Ramus, F. (2021). A Meta-Analytic Review of the Benefit of Spacing out Retrieval Practice Episodes on Retention. *Educational Psychology Review*, 33(3), 959–987. <https://doi.org/10.1007/s10648-020-09572-8>
- Lewis, K., & Kuhfeld, M. (2023). Education’s long COVID: 2022–23 achievement data reveal stalled progress toward pandemic recovery. NWEA.
<https://www.nwea.org/research/publication/educations-long-covid-2022-23-achievement-data-reveal-stalled-progress-toward-pandemic-recovery/>
- McVey, K. P., & Trinidad, J. (2019). Nuance in the Noise: The Complex Reality of Teacher Shortages. Bellwether Education Partners.
- Moats, L. C. (2020). Teaching Reading "Is" Rocket Science: What Expert Teachers of Reading Should Know and Be Able to Do. *American Educator*, 44(2), 4.
- Reutzel, P., Mohr, K. A., & Jones, C. D. (2017). Exploring the relationship between letter recognition and handwriting in early literacy development. *Journal of Early Childhood Literacy*, 19(3), <https://doi.org/10.1177/1468798417728099>
- Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, 50(1), 4-36.
- Skibbe, L. E., Montroy, J. J., Bowles, R. P., & Morrison, F. J. (2019). Self-regulation and the development of literacy and language achievement from preschool through second grade. *Early Childhood Research Quarterly*, 46, 240–251.
<https://doi.org/10.1016/j.ecresq.2018.02.005>
- Squires, K. E., & Wolter, J. A. (2016). The Effects of Orthographic Pattern Intervention on Spelling Performance of Students With Reading Disabilities. *Remedial and Special Education*, 37(6), 357–369. <https://doi.org/10.1177/0741932516631115>
- Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2019). Understanding teacher shortages: An analysis of teacher supply and demand in the United States. *Education Policy Analysis Archives*, 27(35).
- Young, E., & Young, K. (2024a). Student growth in the post-COVID era. Curriculum Associates.
<https://cdn.bfldr.com/LS6J0F7/at/4rqc5wtpxqf85mk4pxj6rm7/ca-2024-summer-research-student-growth-technical-report.pdf>
- Young, K., & Young, E. (2024b). School readiness for young students post-pandemic. Curriculum Associates.



https://cdn.bfldr.com/LS6J0F7/at/p3ttt35vgq78fhvrqq8bwn2x/School_Readiness_for_Young_Students_Post-Pandemic.pdf

DRAFT

Appendix to be added in final version

List of Tables & Figures

Table 1. Sample Sizes by Grade, Year, and Condition for the Unmatched Sample	4
Table 2. Demographic Information for the Matched Samples by Grade and Condition	6
Table 3. Descriptive Statistics for the Matched Samples by Grade and Condition	7
Figure 1. Proficiency Levels for K-1 Spalding and Comparison school students	9
Figure 2. Spring earlyReading Scores for Students Followed from Kindergarten to 1st Grade	10
Figure 3. Spalding Student Growth Compared to National Norms of Expected Growth	11
Figure 4. AASA State Test Proficiency Rates for Grades 3-5 Combined	12
Figure 5. AASA State Test Proficiency Rates for Grades 3-5, Split by Grade	13
Figure 6. AASA State Test Scores for 3rd to 5th Grade Students in Year 2	14
Table 4. Summary of Study Findings Split by Grade	15

Supplementary Materials to be added in final version