EVALUATION OF CURRICULUM ASSOCIATES' LITERACY PORTFOLIO

Quasi-Experimental Effects on Grades K–5 Student Literacy Outcomes

August 2025

Christina LiCalsi, Ph.D.

LiCalsi
EDUCATION EVALUATION CONSULTING

Contents

Background and Purpose

In May 2025, Curriculum Associates contracted with LiCalsi Education Evaluation Consulting to conduct an independent quasi-experimental evaluation estimating the effects of using one or more of their Literacy Portfolio programs on the literacy outcomes of students in Grades K–5. Curriculum Associates' Literacy Portfolio aims to provide intentional instruction for Grades K–5 students through a suite of literacy programs and resources, paired with guidance for teachers on effective instructional use. *The Literacy Portfolio* by *i-Ready®* includes the following programs: adaptive and Diagnostic Literacy Tasks, *Magnetic Foundations™*, *Magnetic Comprehension™*, *Ready® Reading*, *Ready Writing*, *i-Ready Personalized Instruction*, *Phonics for Reading®*, and Teacher Toolbox.

Key Findings

- Students in schools using *The Literacy Portfolio* had significantly higher spring literacy assessment scores, on average, and met Typical Growth and Stretch Growth® targets at rates substantially larger than their matched peers in schools not using *The Literacy Portfolio*. These positive effects were evident across all Grades K–5, with the largest gains in the early grades.
- In schools using *The Literacy Portfolio*, students in Grades K–2 scored significantly higher on the spring assessment phonics subscore, while students in Grades 3–5 scored significantly higher on the comprehension subscore compared with peers in schools not using *The Literacy Portfolio*.
- Literacy Portfolio students outperformed matched peers on the spring literacy assessment regardless of their fall scores, with the largest effects experienced by students who started in the bottom half of the distribution.
- Literacy Portfolio students outperformed matched peers across schools of all socioeconomic levels and baseline literacy performance, with the largest effects in schools in the lowest quartile of income and fall literacy scores.

Study Overview

This evaluation set out to answer the following questions for students in Grades K–5 during the 2023–2024 school year:

EQ1) What is the effect of participating in one or more of *The Literacy Portfolio* programs on (a) students' overall spring Diagnostic assessment score, (b) students' spring phonics Diagnostic subscale score (Grades K–2 only), (c) students' spring comprehension Diagnostic subscale score (Grades 3–5 only), (d) the likelihood that students met their Typical Growth target, and (e) the likelihood that students met their Stretch Growth target?

EQ2) Do the effects of participating in a Literacy Portfolio program differ depending on students' (a) race, (b) gender, (c) English Learner status, (d) disability, (e) grade, or (f) fall Diagnostic score quartile?

Because student race, gender, English Learner, and disability data were missing for a large portion of students, it was not possible to estimate differential effects by these characteristics as planned in EQ2a through EQ2d. Instead, subgroup effects were estimated by student grade and fall Diagnostic score quartile (i.e., EQ2e, EQ2f), schools' ZIP code median household income quartile, and schools' average fall Diagnostic test score quartile.

Methods

This evaluation uses a matched comparison design to estimate the effect of Literacy Portfolio program usage by pairing each student in a school that used Literacy Portfolio programs with a similar student in a school that did not, and comparing spring literacy Diagnostic assessment outcomes for the two groups. Student performance was measured with Curriculum Associates' fall and spring *i-Ready Diagnostic* for Reading assessment. The Diagnostic is an online, adaptive, and criterion-referenced assessment of student learning for reading in Grades K–8. Students who take the *i-Ready Diagnostic* receive a scale score that reflects their test performance. Scores are on the same scale across grades to allow for comparisons to be made over time. By matching students who used a Literacy Portfolio program to those who did not on key pre-intervention characteristics, including fall Diagnostic scores, this design can reduce selection bias and come closer to approximating the causal inference of random assignment.

The strength of a matched comparison design rests on two key factors: a large, credible pool of potential comparison individuals, and rich pre-intervention data. A credible comparison pool must include individuals who closely resemble the intervention group on covariates linked to the outcomes of interest (i.e., strong common support). When common support is imperfect—for example, if the intervention group includes test scores outside the comparison group's range—individuals outside of the overlapping range are dropped, since no valid counterfactual exists. As a result, estimates apply only to the population within the common-support region, not the entire intervention group. Credibility is also enhanced when the comparison group is observed over the same time frame as the intervention group but lacked access to the intervention. Otherwise, individuals who had access but declined participation have demonstrated that they likely differ in unobserved ways (e.g., motivation, parental engagement) that bias estimates. Similarly, if comparison individuals interact with treated individuals, spillover effects can dilute estimated impacts.

The richness of pre-intervention data is determined primarily by how well it predicts the outcome of interest. Because prior performance strongly predicts future performance, baseline outcome measures are typically the most valuable, especially when highly correlated with post-intervention outcomes. Other important pre-intervention factors often include prior trends, demographics, and contextual variables. Finally, variables that predict intervention assignment are useful for ensuring treated individuals are matched with peers from similar circumstances.

This evaluation features a strong matched comparison design that meets all of these standards. The remainder of this section will detail the data, sample selection, and design elements, highlighting where design decisions strengthen confidence in the estimated effects.

Data

Data for the evaluation came from four files: (1) a school-level file with 77 schools across three districts that purchased *The Literacy Portfolio* to use with their students, (2) a school-level file with 4,718 schools that used the *i-Ready Diagnostic* for Reading assessment with students but had not purchased *The Literacy Portfolio*, and (3–4) student-level files for students in these schools (24,259 students in Literacy Portfolio schools and 237,483 students in non-Literacy Portfolio schools) during the 2023–2024 school year. The school-level files included identifying information (i.e., school name, district, state), average fall, winter, and spring Diagnostic scores by grade (overall and by subscore), and National Center for Education Statistics (NCES) demographic and enrollment data. Student-level files included identifiers (i.e., student ID, school ID, grade), demographic characteristics, and fall, winter, and spring Diagnostic scores (overall and by subscore).

Sample Selection and Design

To be included in the evaluation, students needed to have both a fall and spring Diagnostic score included in the data files. There were 21,781 students in 77 schools that purchased *The Literacy Portfolio*, and 216,579 potential matched comparison students in 791 schools that had not purchased *The Literacy Portfolio*, who met these criteria. To ensure that the matched comparison group attended schools that were similar to Literacy Portfolio schools, only students attending schools that had the necessary NCES data to generate propensity scores were included. This last condition did not change the sample of students who did not have access to *The Literacy Portfolio* but reduced the sample of Literacy Portfolio students to 20,737 students in 69 schools.

The design used in this evaluation first matched students' grade and fall Diagnostic score exactly. Within each grade, the range of fall Diagnostic scores was both lower and higher in the potential comparison pool than for Literacy Portfolio students, and exact matches were available for all Literacy Portfolio students. From among the set of same-grade, same-score, potential comparison students, one-to-one matches were made, without replacement, to the comparison student with the closest propensity score (i.e., nearest neighbor), constrained by a 0.2 standard deviation (SD) caliper on the logit of the propensity score to ensure that students who had access to *The Literacy Portfolio* were matched to comparison students in schools that were not starkly different in terms of overall income, performance, and size.

Choosing the right propensity score model requires balancing the inclusion of enough covariates to ensure the intervention and comparison groups are similar on key characteristics without including so many covariates that it becomes overly difficult to find suitable matches, reducing common support and limiting the sample available for analysis. To achieve this balance, propensity scores were constructed using the average fall Diagnostic score, by grade, in the student's school, the median household income in the ZIP code where the student's school was located, and the overall school enrollment. School racial and ethnic makeup and locale could not be included as covariates in the propensity score because the intervention schools came

exclusively from three districts, drastically reducing the variation on these measures within the intervention group and the number of possible matches to draw from in the comparison group. Literacy Portfolio students who lacked a comparison within the caliper were excluded[1], enforcing common support. This "exact-then-propensity score" approach concentrates bias reduction where it matters most, ensuring Literacy Portfolio users and comparison students of the same grade started at the same fall baseline—by far the strongest predictor of spring performance, and fine-tunes residual differences on other factors.

Analytic Sample

Table 1 summarizes student characteristics for the 20,635 comparison students, the 20,635 students who used a Learning Portfolio program, and the full sample. Among the students in schools using *The Literacy Portfolio*, Grade K students made up a smaller percentage (11.71) of students than students in Grades 1–5 (16.12 to 18.79), a pattern that is replicated among comparison students because Literacy Portfolio users are matched to comparison students with the same fall Diagnostic score in the same grade. It is difficult to make any true comparisons regarding the demographic characteristics of comparison students and students in Literacy Portfolio schools because of the high rate of missing demographic data among Literacy Portfolio schools and the fact that all demographic data was missing for all schools in one of the three districts using *The Literacy Portfolio*. As such, student demographic data are not included in any regression models.

Table 1

Student Baseline Characteristics by Literacy Portfolio Use

Baseline Characteristic	Comparison Students		Literacy Portfolio Users		Full Sample		Standardized Difference
	N	%	*N*	%	*N*	%	
Grade							
K	2,416	11.71	2,416	11.71	4,832	11.71	--
1	3,326	16.12	3,326	16.12	6,652	16.12	--
2	3,878	18.79	3,878	18.79	7,756	18.79	--
3	3,793	18.38	3,793	18.38	7,586	18.38	--
4	3,725	18.05	3,725	18.05	7,550	18.05	--
5	3,497	16.95	3,497	16.95	6,994	16.95	--
Gender							
Male	9,492	46.0	3,669	17.8	13,161	31.89	-0.635
Female	9,445	45.8	3,532	17.1	12,977	31.44	-0.649
Missing	1,698	8.2	13,434	65.1	15,132	36.67	1.462
Race							
Black	1,864	9.0	2,045	9.9	3,909	9.47	0.030
White	10,444	50.6	3,632	17.6	14,076	34.11	-0.743

[1]102 students in the intervention group were unable to be matched to a comparison student within the 0.2 SD propensity score caliper. This represented 0.49% of the 20,737 Literacy Portfolio students for whom matches were attempted.

Asian	942	4.6	90	0.4	1,032	2.50	-0.267
American Indian/Alaskan	215	1.0	374	1.8	589	1.43	0.065
Native Hawaiian/ Pacific Islander	71	0.3	0	0.0	71	0.17	-0.083
Two+ Races	1,384	6.7	1,062	5.1	2,446	5.93	-0.066
Missing	5,704	27.6	13,432	65.1	19,136	46.37	0.810
Disabled							
Yes	1,773	8.59	1,486	7.20	3,259	7.90	-0.052
No	9,883	47.89	5,712	27.68	15,595	37.79	-0.426
Missing	8,979	43.51	13,437	65.12	22,416	54.32	0.444
English Learner							
Yes	1,331	6.45	1,504	7.29	2,835	6.87	0.033
No	9,226	44.71	19,126	92.69	28,352	68.70	1.209
Missing	10,078	48.84	5	0.02	10,083	24.43	-1.380

Table 2 presents school characteristics for the schools attended by comparison students, Literacy Portfolio users, and the full sample. While Literacy Portfolio users and matched comparison students had the same average fall Diagnostic scores, their peers in the same grade at Literacy Portfolio schools scored slightly higher—by about two to five points across grades. Literacy Portfolio schools were also somewhat larger on average (461 versus 444 students), located in neighborhoods with slightly higher average incomes ($76,211 versus $73,507), more likely to be in suburbs (44.7% versus. 30.4%) and towns (17.0% versus 11.5%), and served a substantially lower share of Hispanic students (10.1% versus 22.2%).

Table 2

Baseline Characteristics of the Schools That Students Attended by Literacy Portfolio Use

Baseline Character-istic	Comparison Students		Literacy Portfolio Users		Full Sample			Standardized Difference
	Mean	SD	Mean	SD	Mean	SD	*N*	
Average Fall Score								
K	346.49	9.87	348.50	10.36	347.50	10.17	4,832	-0.199
1	396.53	14.74	401.67	15.77	399.10	15.48	6,652	-0.337
2	448.77	20.47	453.74	20.14	451.25	20.46	7,756	-0.245
3	492.45	22.94	495.50	21.48	493.97	22.27	7,586	-0.137
4	523.12	22.96	526.57	23.41	524.84	23.25	7,450	-0.149
5	550.47	23.47	554.81	24.02	552.64	6,994	6,994	-0.183
Enrollment	443.80	180.77	460.50	163.18	452.16	172.40	41,270	-0.097
Median Household Income	$73,507	$19,028	$76,211	$13,032	$74,859	$16,363	41,270	-0.166
Percentage Black	9.45	14.87	10.90	25.42	10.17	20.84	41,270	-0.07
Percentage Hispanic	22.20	19.96	10.11	4.51	16.15	15.68	41,270	0.84

Days between Assessment	241.62	19.03	249.62	10.61	245.62	15.92	41,270	-0.519
	N	%	*N*	%	*N*	%		
Locale								
City	7,805	37.8	5,820	28.2	13,625	33.01	--	0.206
Suburban	6,264	30.4	2,079	44.7	6,267	15.19	--	-0.300
Town	2,378	11.5	9,224	17.0	15,488	37.53	--	-0.157
Rural	4,188	20.3	3,512	10.1	5,890	14.27	--	0.288

Analysis and Outcome Measures

The evaluation examined the effect of Literacy Portfolio program use on three main spring literacy outcomes: (1) students' average overall spring Diagnostic assessment scores, (2) the percentage of students who met Typical Growth targets, and (3) the percentage of students who met Stretch Growth targets. Typical Growth and Stretch Growth targets are targets for spring assessment scores that are based on students' fall scores and informed by the distribution of fall-to-spring student score gains. The evaluation also examined the effects of *The Literacy Portfolio* on two important subscales: phonics (for Grades K–2 only) and comprehension (for Grades 3–5 only).

The effects of Literacy Portfolio program use (EQ1) were estimated using a series of linear regressions in which student spring Diagnostic score outcomes were regressed on an indicator for Literacy Portfolio use, controlling for fall Diagnostic score, school-level characteristics, and the exact number of days between fall and spring Diagnostic tests. The model also included grade-fixed effects so students are only compared within grades, and clustered standard errors at the school level to account for Diagnostic score correlations between students in the same school.

The equation used is as follows:

$$(1)\ Yij = \beta 0 + \beta 1 \mathrm{LP}j + \boldsymbol{Xij}\beta StudentCov + \boldsymbol{Wj}\beta SchoolCov + GradeFE + vj + eij$$

Wherein:

Yij measures the outcome of interest
$\boldsymbol{X}ij$ is a vector of student-level covariates
$\boldsymbol{Wj}$ is a vector of school-level covariates
GradeFE is a vector of dummy indicators for Grades K–5
vj measures the school-level residual
eij measures the student-level residual

β1 provides a covariate-adjusted estimate of the effect of being in a school that used Literacy Portfolio programs on students' average spring literacy outcome measures. In other words, it represents the average difference in outcomes between students in schools using *The Literacy Portfolio* and students with the same fall Diagnostic scores but who did not have the opportunity to use a Literacy Portfolio program because their school did not choose to purchase

The Literacy Portfolio, after controlling for the covariates in the model. Subgroup analyses were conducted using the same models but for subsets of the total analytic sample constructed based on student grade, student fall Diagnostic score quartile, schools' ZIP code median household income quartile, and schools' average fall test score quartile. To aid interpretation, both unstandardized and standardized effect estimates are reported where appropriate. Unstandardized coefficients provide results in the natural units of the outcome (e.g., test score points), which are intuitive and meaningful for practitioners and policymakers. Standardized effect sizes, expressed in SD units, facilitate comparisons across different grades, outcomes, and measures that are on different scales. Reporting both estimates offers a fuller picture of program impacts by combining real-world interpretability with cross-context comparability.

Findings

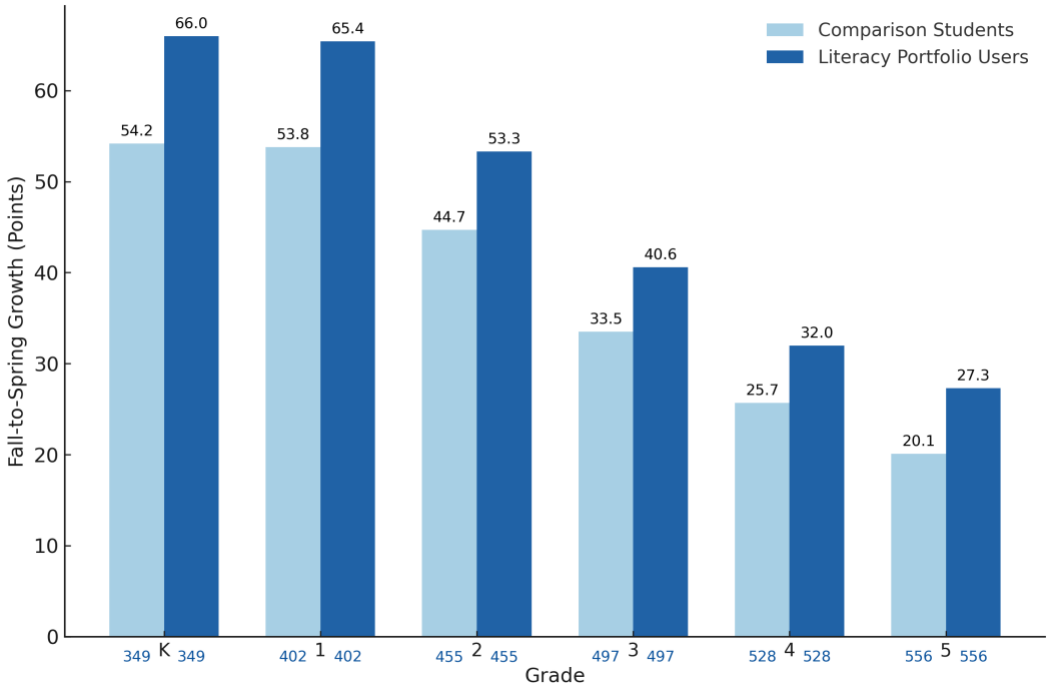
This section first presents descriptive spring literacy outcomes for the full analytic sample. Descriptive findings are followed by regression estimates of the effects of Literacy Portfolio usage on spring outcomes controlling for remaining covariate differences (EQ1). Lastly, effect estimates are presented for student and school subgroups (EQ2).

Students Who Used Literacy Portfolio Programs Outperformed Comparison Students by a Substantial Margin

Descriptive comparisons of fall-to-spring Diagnostic score growth, without adjusting for covariates, showed that Literacy Portfolio students demonstrated greater fall-to-spring score gains than comparison students at every grade level (see Figure 1). Overall, average score gains for both groups of students were largest in Grades K and 1 and then decreased in a stepwise fashion through Grade 5, at which point the magnitude of average score gains was less than half what it was in the earliest grades. This pattern is important to keep in mind when comparing unstandardized effect estimates because *the same size effect, as measured in spring Diagnostic score points, represents a larger improvement proportionally for later grades*. For example, although Grade K students who used a Literacy Portfolio program outscored the matched comparison group by 11.8 points, this represents a fall-to-spring score gain for Literacy Portfolio students that is 17.9 percent larger (11.8/54.2) than was experienced by the matched comparison students. Meanwhile, Grade 5 students who used a Literacy Portfolio program outscored the matched comparison group by only 7.2 points, yet this represents a fall-to-spring score gain that is 26.4 percent larger (7.2/20.1) than was experienced by matched comparison students. Standardized effect sizes will account for these differences and aid in interpretation.

Figure 1

Fall-to-Spring Growth in Diagnostic Scores by Grade and Literacy Portfolio Use

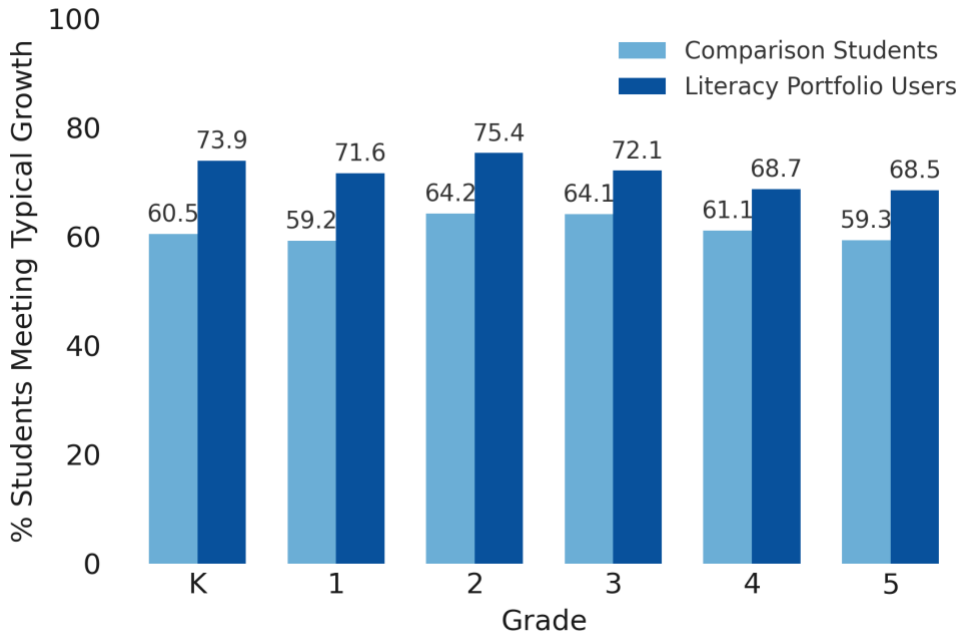


Note: Bars represent mean fall-to-spring growth in Diagnostic scores with growth values shown above bars. Numbers below the *x*-axis are mean fall baseline scores for each group.

Not only were average score gains greater for Literacy Portfolio users, across all grades, more Literacy Portfolio users met Typical Growth targets (see Figure 2) and Stretch Growth targets (see Figure 3). Grade K students in particular experienced the largest increases in meeting Typical Growth and Stretch Growth targets, with differences representing 22.0 percent and 45.7 percent increases in the proportion of students who met Typical Growth and Stretch Growth targets, respectively.

Figure 2

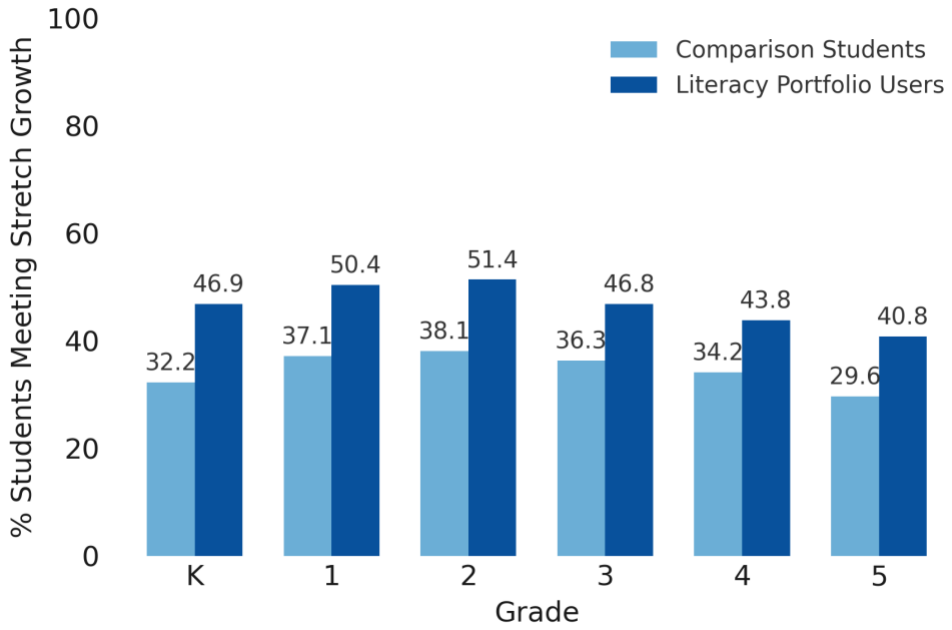
Percentage of Students Meeting Typical Growth Targets by Grade and Literacy Portfolio Use



Note: Bars represent the percentage of students meeting Typical Growth targets.

Figure 3

Percentage of Students Meeting Stretch Growth Targets by Grade and Literacy Portfolio Use



Note: Bars represent the percentage of students meeting Stretch Growth targets.

Students Who Used Literacy Portfolio Programs Outperformed Comparison Students on All Outcomes Even after Controlling for Differences in the Schools They Attended

Results from regression analyses controlling for any differences between test-taking times and school covariates also found that students who used *The Literacy Portfolio* outperformed their peers across all outcomes (see Table 3). On average, Literacy Portfolio students scored nearly seven points higher on the spring Diagnostic assessment (*b* = 6.963, SE = 0.889), representing a difference of approximately 14 percent of an SD. Literacy Portfolio users were also eight percentage points more likely to meet Typical Growth targets (*b* = 0.080, SE = 0.009) and 9.5 percentage points more likely to meet Stretch Growth targets (*b* = 0.094, SE = 0.010) compared with matched comparison students. All effects are of a substantial magnitude and are statistically significant at the 0.01 level, providing strong evidence that Literacy Portfolio use results in improved literacy outcomes for students. Again, even after controlling for additional baseline school-level differences, the effect of Literacy Portfolio usage on meeting Stretch Growth targets is particularly large in magnitude, representing a 26 percent increase in the number of students compared to the comparison group.

Table 3

Effect Estimates of Literacy Portfolio Use on Spring Diagnostic Outcomes

Predictor	Spring Score Points		Spring Score SD Units		Met Typical Growth		Met Stretch Growth		*N*
	B	SE	*β*	SE	*B*	SE	*B*	SE	
Literacy Portfolio Use	6.963**	0.889	0.138**	0.019	0.080**	0.009	0.094**	0.010	41,270

Note: *B* = unstandardized coefficient; SE = standard error; *β* = standardized coefficient. * $p < .05$, ** $p < .01$.

Literacy Portfolio students in Grades K–2 also performed more than nine points better on the *i-Ready* spring Diagnostic assessment phonics subscale (*b* = 9.458, SE = 1.505), and nearly six points better on the comprehension subscale in Grades 3–5 (*b* = 5.768, SE = 0.911). These findings suggest that use of Literacy Portfolio programs has a substantive positive effect on both early foundational skills and later reading comprehension (see Table 4).

Table 4

Effect Estimates of Literacy Portfolio Use on Spring Phonics and Comprehension Subscale Scores

	Grades K–2				
Predictor	Phonics Score Points		Phonics Score SD Units		*N*
	B	SE	*β*	SE	
Literacy Portfolio Use	9.458**	1.505	0.163**	0.026	19,240
	Grades 3–5				

	Comprehension Score Points		Comprehension Score SD Units		*N*
	B	SE	*β*	SE	
Literacy Portfolio Use	5.768**	0.911	0.095**	0.015	22,030

Note: *B* = unstandardized coefficient; SE = standard error; *β* = standardized coefficient. * $p < .05$, ** $p < .01$.

Students of All Ages and Baseline Literacy Levels Benefited from Using Literacy Portfolio Programs

Regression analyses disaggregated by grade found that Literacy Portfolio program use was consistently associated with positive outcomes across grade levels, with the largest effects found in the early grades (see Table 5). Grade K students in particular experienced large positive effects with spring assessment scores that were almost a quarter of an SD larger than that of matched comparison students. Students in schools using *The Literacy Portfolio* were also significantly more likely to meet both Typical Growth and Stretch Growth targets in every grade, with effects ranging from 5.5 percentage points (*B* = 0.055, SE = 0.017) to 12.6 percentage points (*B* = 0.126 SE = 0.020) higher than the matched comparison group. These findings suggest that Literacy Portfolio use was broadly beneficial for students of all ages.

Table 5

Effect Estimates of Literacy Portfolio Use on Spring Diagnostic Outcomes by Grade

Grade	Spring Score Points		Spring Score SD Units		Met Typical Growth		Met Stretch Growth		*N*
	B	SE	*β*	SE	*B*	SE	*B*	SE	
K	9.103**	2.044	0.240**	0.054	0.112**	0.025	0.123**	0.028	4,832
1	8.631**	1.751	0.177**	0.036	0.092**	0.019	0.097**	0.021	6,652
2	8.479**	1.736	0.163**	0.033	0.102**	0.017	0.126**	0.020	7,756
3	5.402**	1.281	0.099**	0.024	0.060**	0.017	0.083**	0.017	7,586
4	4.872**	1.154	0.085**	0.020	0.055**	0.017	0.078**	0.018	7,450
5	6.588**	1.100	0.115**	0.019	0.082**	0.018	0.097**	0.017	6,994

Note: *B* = unstandardized coefficient; SE = standard error; *β* = standardized coefficient. * $p < .05$, ** $p < .01$.

Regression analyses disaggregated by fall Diagnostic score quartile indicate that Literacy Portfolio program use has positive effects for students with all levels of baseline literacy scores (see Table 6). Effects ranged from approximately 4.8 points for students in the highest quartile (*B* = 4.785, SE = 0.792) to nearly eight points for students with fall scores in the lowest quartile (*B* = 7.901, SE = 1.646). In addition, Literacy Portfolio students were significantly more likely to meet both Typical Growth and Stretch Growth targets in every quartile. Students in schools using *The Literacy Portfolio* met Typical Growth standards at rates between 6.7 and 9.5 percentage points higher (*B* = 0.067, SE = 0.016 and *B* = 0.095, SE = 0.013) than students in schools that did not use *The Literacy Portfolio*. Similarly, Literacy Portfolio students met Stretch

Growth standards at rates between 8.5 and 11.0 percentage points higher (*B* = 0.085, SE = 0.013 and *B* = 0.110, SE = 0.015) compared to matched students who did not use *The Literacy Portfolio*. These results suggest *The Literacy Portfolio* produced consistent and large benefits for students across the performance distribution.

Table 6

Effect Estimates of Literacy Portfolio Use on Spring Diagnostic Outcomes by Fall Score Quartile

Fall Score Quartile	Spring Score Points		Spring Score SD Units		Met Typical Growth		Met Stretch Growth		*N*
	B	SE	*β*	SE	*B*	SE	*B*	SE	
1	7.901**	1.646	0.154**	0.032	0.067**	0.016	0.094**	0.015	10,428
2	7.860**	1.028	0.155**	0.021	0.090**	0.013	0.110**	0.015	10,398
3	6.684**	0.825	0.133**	0.017	0.095**	0.013	0.099**	0.013	10,358
4	4.785**	0.792	0.102**	0.017	0.066**	0.013	0.085**	0.013	10,086

Note: *B* = unstandardized coefficient; SE = standard error; *β* = standardized coefficient. * $p < .05$, ** $p < .01$.

Literacy Portfolio Program Use Had Positive Effects on Students' Spring Literacy Outcomes across Diverse School Contexts

The benefits of *The Literacy Portfolio* extended across schools serving different income levels (see Table 7). Across all income quartiles, students in schools using *The Literacy Portfolio* scored substantially higher on the spring Diagnostic assessment and were significantly more likely to meet both Typical Growth and Stretch Growth targets. Students in schools in the lowest quartile of income experienced the largest effects with spring scores 26.5 percent of an SD higher (*B* = 0.265, SE = 0.074) than matched comparison students and rates of meeting Typical Growth and Stretch Growth targets that were 15.9 and 15.0 percentage points higher than comparison students, respectively (*B* = 0.159, SE = 0.074 and *B* = 0.150, SE = 0.040).

Table 7

Effect Estimates of Literacy Portfolio Use on Spring Diagnostic Outcomes by School Median Household Income Quartile

School Income Quartile	Spring Score Points		Spring Score SD Units		Met Typical Growth		Met Stretch Growth		*N*
	B	SE	*β*	SE	*B*	SE	*B*	SE	
1	13.708**	3.692	0.265**	0.074	0.159**	0.043	0.150**	0.040	10,390
2	5.443**	1.788	0.103**	0.037	0.070**	0.026	0.110**	0.024	13,414
3	10.617**	3.114	0.218**	0.069	0.061**	0.023	0.137**	0.020	7,273
4	6.746**	1.130	0.142**	0.021	0.087**	0.016	0.100**	0.019	10,193

Note: *B* = unstandardized coefficient; SE = standard error; *β* = standardized coefficient. * $p < .05$, ** $p < .01$.

When results were disaggregated by school average fall Diagnostic score quartile, Literacy Portfolio use had positive effects on spring Diagnostic scores and the likelihood of meeting Typical Growth and Stretch Growth targets across the distribution. The largest effects were experienced by students in schools with the lowest average fall Diagnostic score performance, however, students in the bottom quartile of schools had spring Diagnostic scores that were 18.5 percent of an SD larger than comparison students (B = 0.185, SE = 0.038) and were more than nine percentage points more likely to meet both Typical Growth and Stretch Growth targets (B = 0.093, SE = 0.018 and B = 0.099, SE = 0.021). These results suggest that although students in all schools benefit from using *The Literacy Portfolio*, students in lower-performing schools may benefit the most.

Table 8

Effect Estimates of Literacy Portfolio Use on Spring Diagnostic Outcomes by School Average Fall Diagnostic Score Quartile

School Fall Score Quartile	Spring Score Points		Spring Score SD Units		Met Typical Growth		Met Stretch Growth		*N*
	B	SE	*β*	SE	*B*	SE	*B*	SE	
1	8.062**	1.622	0.185**	0.038	0.093**	0.018	0.099**	0.021	10,439
2	8.209**	1.634	0.156**	0.031	0.092**	0.017	0.117**	0.018	10,291
3	5.567**	1.175	0.099**	0.021	0.060**	0.016	0.085**	0.016	10,306
4	5.701**	0.837	0.099**	0.015	0.077**	0.016	0.090**	0.015	10,234

Note: B = unstandardized coefficient; SE = standard error; β = standardized coefficient. * $p < .05$, ** $p < .01$.

Discussion

This quasi-experimental matched comparison evaluation found consistent, highly statistically significant positive effects on spring literacy outcomes for students attending a school that used *The Literacy Portfolio*. Effects were positive and statistically significant at the < 0.01 level for every outcome and across every subgroup of students. Across all grade levels, students using *The Literacy Portfolio* had higher spring Diagnostic assessment scores and met Typical Growth and Stretch Growth targets at higher rates than matched comparison students with the same fall Diagnostic scores. The overall standardized effect size was 0.138 SD units, with students in the earliest grades experiencing the largest effects—0.240, 0.177, and 0.163 SDs for Grades K, 1, and 2 students, respectively. Estimates also suggest that attending a school using *The Literacy Portfolio* may be particularly beneficial for students who are struggling the most and in the most challenging circumstances. Effects were largest for students who began the year with the lowest literacy scores—0.155 SDs, and students who attended schools in the lowest-income neighborhoods—0.265 and with the poorest average literacy levels—0.185 SDs. These effect sizes are of a similar magnitude to other high-quality, evidence-based literacy programs and are relatively large for a whole school program (Alrawashdeh et al., 2024; Cortes et al., 2025; Neitzel et al., 2022; Wanzek et al., 2018).

Overall, the matched comparison design provides convincing evidence that students experience positive effects when their schools use *The Literacy Portfolio*. Students in both *The Literacy Portfolio* group and the comparison group took the same assessment and were matched exactly on fall baseline scores and within a 0.20 caliper on propensity scores that included schools' enrollment, median income, and average fall assessment scores. Further, matches were made for more than 99.5 percent of students in Literacy Portfolio schools who met the criteria for inclusion, providing confidence that estimates are representative and not due to limited common support.

This evaluation has two important limitations to consider. The main limitation is that all 69 Literacy Portfolio schools were drawn from only three districts, which constrains external validity and makes findings sensitive to unobserved district-level changes; with so few districts, unobservable changes unrelated to Literacy Portfolio use in any one district could disproportionately bias the overall effect. The small number of districts also constrained matching: the lack of demographic variability stemming from the lack of district variation meant that schools were not able to be matched on demographic characteristics districts, state-fixed effects were not able to be used, limiting the ability to account for state-specific policy or historical factors, and propensity scores were limited in the covariates they could make use of. Further, one district using *The Literacy Portfolio* did not report student-level demographics, and as a result, student-level covariates could not be included and student subgroup effects could not be examined.

Another important limitation is that it is not possible to determine from the data anything about how or the extent to which schools were using Literacy Portfolio programs. The data did not include any information about which Literacy Portfolio programs were being used, the extent to which teachers were using the programs as intended and in accordance with guidance, or the frequency or intensity of use. It was also not possible to determine if all students in a school, for whom data was available, actually used any of the programs. Although *The Literacy Portfolio* is meant to be used broadly across grades and for all students, it is possible that in some schools or classrooms this did not occur. These data limitations mean that the estimates presented should be interpreted as intent-to-treat estimates and therefore may be a lower-bound estimate of the true effect of *The Literacy Portfolio* when used as intended for all students.

ESSA Evidence

Although this evaluation employed a quasi-experimental design, based on guidance provided by The Institute for Education Science, it is likely to meet Tier 3 evidence standards (not Tier 2). This evaluation's design meets many of the criteria necessary to qualify for Tier 2 evidence standards; most importantly, although not randomly assigned, students in schools using *The Literacy Portfolio* are perfectly matched at baseline to students with the same fall assessment score but in a school that does not use *The Literacy Portfolio*. Further, all effects are positive and statistically significant. However, to meet Tier 2 standards, a quasi-experimental study must be prospective—or, if retrospective, must emulate a prospective design by including all intervention cases with baseline data and matching comparisons using baseline measures only.

Individuals are then dropped after matching if they do not have an outcome measure and baseline equivalence must hold for the analytic sample. Additionally, to meet Tier 2 standards, studies should be large and from multiple sites. Because the Literacy Portfolio schools and the schools that potential comparison students attended differed substantially at baseline, nesting was highly unbalanced, the intervention sample came from only three, relatively homogeneous districts, and approximately 10 percent of students in each group were missing spring scores, matching was restricted to students preselected to have both fall and spring scores. This choice strengthens the design by ensuring baseline equivalence on the fall Diagnostic within the analytic sample.

Next Steps

The results of this evaluation provide promising evidence that when schools use *The Literacy Portfolio*, their students' literacy performance improves more than it would doing business as usual. Further, the consistently positive effects across diverse students and schools should garner confidence among school and district leaders that using *The Literacy Portfolio* will help *their* students in particular. To build on the evidence established in this evaluation, future research should aim to accomplish the following:

- Include schools from a larger number of districts, ideally districts that are large and diverse. More, and more diverse, districts will reduce concerns that estimates are biased due to unobservable district or state contextual factors. It will also increase the external validity of estimates and instill confidence that when school leaders implement *The Literacy Portfolio*, their students will benefit.
- Work with districts to provide student-level demographic data so analyses can control for differences in student characteristics and student subgroup analyses can explore differential effects and equitable outcomes for students of diverse backgrounds and with diverse needs.
- Include a process evaluation component to collect data on how teachers and schools are using *The Literacy Portfolio*, with whom, how much, and why. Qualitative data collection, surveys, and measures of usage and implementation for various portfolio elements will help provide contextualization and allow for continuous improvement moving forward.

References

Alrawashdeh, G. S., Fyffe, S., Azevedo, R. F., & Castillo, N. M. (2024). Exploring the impact of personalized and adaptive learning technologies on reading literacy: A global meta-analysis. *Educational Research Review*, *42*, 100587.

Cortes, K. E., Kortecamp, K., Loeb, S., & Robinson, C. D. (2025). A scalable approach to high-impact tutoring for young readers. *Learning and Instruction*, *95*, 102021.

Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2022). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly*, *57*(1), 149–179.

Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities, 51*, 612–624.