ATTLEBORO PUBLIC SCHOOLS PILOT DATA COLLECTION PROJECT

EVALUATION REPORT

June 2025

Prepared for

Benchmark Education
Lindsay Marczak
Lmarczak@benchmarkeducation.com

Prepared by

Margaret Beam
Beam Consulting LLC
beamconsultingpdx@gmail.com

CONTENTS

EXHIBITS

OVERVIEW

DISTRICT CHARACTERISTICS

Attleboro Public Schools is a public school district located in Attleboro, Massachusetts. The suburban district includes an early learning center, five elementary, three middle, and three high schools, and serves approximately 6,200 students in pre-kindergarten through Grade 12. Approximately 20% of enrolled students identify as Hispanic or Latino, 40% of families are economically disadvantaged, and 19% are designated as special education. Approximately 10% of students are classified as English language learners, and 19% have a first language other than English. According to state test scores, 41% of students are at least proficient in math and 39% in reading. On the spring 2024 Massachusetts Comprehensive Assessment System (MCAS), 33% of students in Grades 3-8 districtwide were proficient in reading and 41% in math compared to 39% and 41% statewide for reading and math, respectively.[1]

IMPLEMENTATION OVERVIEW

In spring 2025, Attleboro Public Schools piloted programs from Benchmark Education Company with multilingual learners, newcomers, and students enrolled in English Language Development (ELD) programs, including its Express! program, a K-6 ELD program that promotes language learning. Attleboro Public Schools implemented Express! with 206 eligible students in Grades K-4 in three of the district's five schools.

Benchmark Education collaborated with Attleboro Public Schools to explore the implementation and influence of Express! on student learning outcomes. Attleboro Public Schools provided Benchmark Education with student demographic data and Renaissance Star and DIBELS assessment data to track student progress and growth in reading and language skills and assess how Express! supports ELD students' literacy development.

EVALUATION OVERVIEW

The evaluation used a two-step design to answer the following four questions:

1. Do ELD students receiving Express! instruction demonstrate statistically significant gains on standardized reading tests?
2. Do gains on standardized reading tests of ELD students receiving Express! instruction differ by starting performance level?
3. Do ELD students receiving Express! instruction demonstrate statistically greater growth or higher performance on standardized reading tests than an equivalent group of students not using the program?
4. Do reading achievement gains vary based on student demographic characteristics?

Evaluation Questions 1 and 3 were addressed using a pre-post design in which all Express! students were assessed on their pre- and post-assessment scores. Evaluation Question 2 was addressed using a matched-comparison design with a subsample of Express! and comparison students who were matched one-to-one on key demographic characteristics and baseline assessment scores using propensity scores.

[1] https://profiles.doe.mass.edu/analysis/nextgenmcas.aspx?orgcode=00160000&orgtypecode=5&compareorg=&

EVALUATION DESIGN

This evaluation was initially designed as a quasi-experimental study in which an intervention group of Benchmark Education's Express! students (ELD students in three elementary schools) would be compared to a group of similar non-Express! students (ELD students in the district's remaining two elementary schools). However, post-implementation analyses determined that the comparison sample was not equivalent to the Express! sample on key demographic characteristics or baseline assessment scores (see Appendix A for detailed description of samples and rationale for shift in design). To address this challenge, the evaluation modified the design to a two-step approach:

1. Pre-post design in which all Express! students were assessed on their pre- and post-assessment score gains
2. Matched-comparison design with a subsample of Express! and comparison students who were one-to-one matched on key demographic characteristics and baseline assessment scores using propensity score next-neighbor matching.

PARTICIPANTS

The Express! sample includes 206 K-4 students who participated in the program in spring 2025 in three schools. The demographic characteristics of the combined-grade samples are presented in Exhibit 1.

Exhibit 1. Express! Sample Characteristics

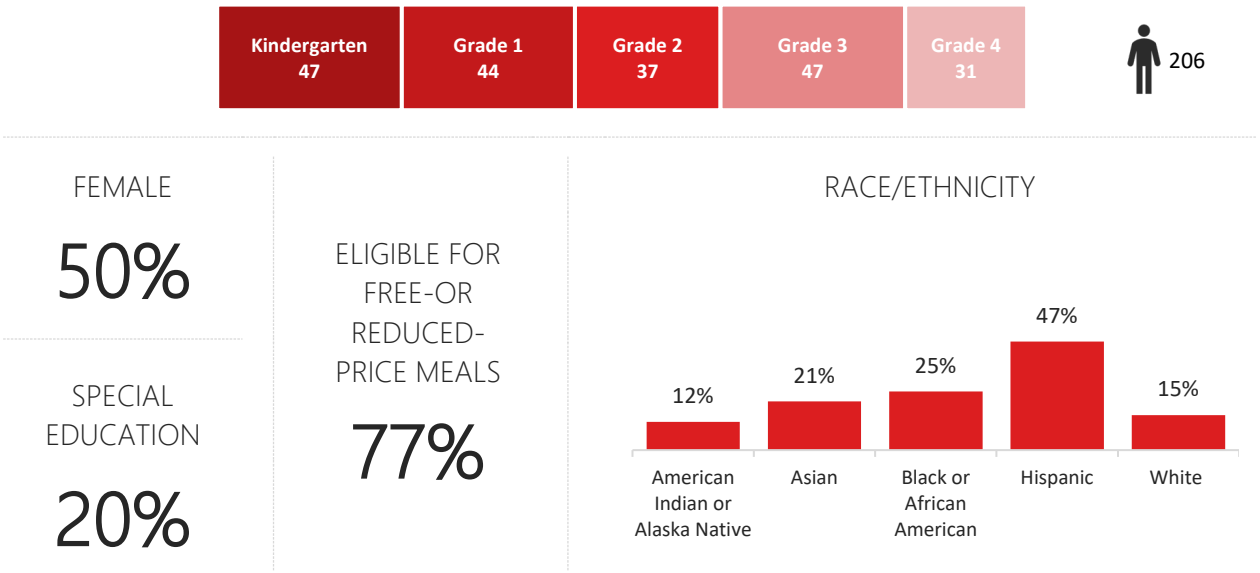
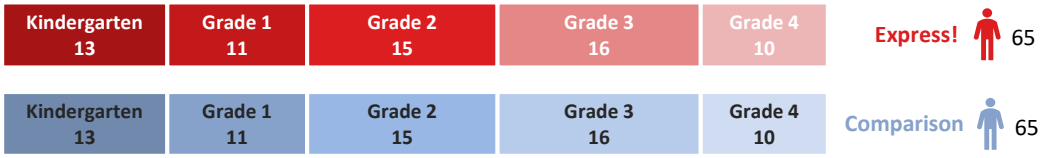


Exhibit 2 presents demographic characteristics for the matched subsample that includes 65 K-4 students who participated in Express! in spring 2025 and 65 matched comparison students from the district's two other schools.

Exhibit 2. Matched Express! and Comparison Subsample Characteristics



FEMALE

43% 42%

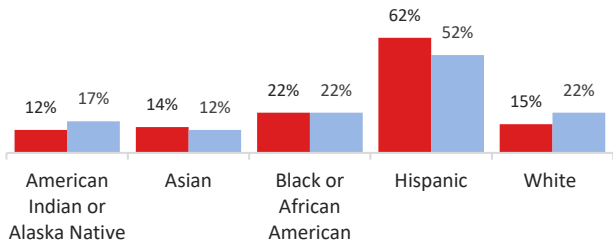
SPECIAL EDUCATION

17% 18%

ELIGIBLE FOR FREE-OR REDUCED-PRICE MEALS

85% 92%

RACE/ETHNICITY



MEASURES

Attleboro Public Schools administered three assessments to its English learner students in kindergarten through Grade 4 in winter and spring of the 2024–2025 school year:

- Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and
- Renaissance Star.

Both scale scores and normal curve equivalent (NCE) scores from these assessments are used in this evaluation. Whereas scale scores reflect absolute academic growth, NCE scores indicate how a student's performance ranks relative to a national norm group (e.g., students of the same age or grade level across the nation). Thus it is possible to see significant scale score growth but nonsignificant NCE growth if their peers nationwide are progressing at a similar rate. NCE scores can be analyzed across grade levels; however, unless vertically scaled, scale scores cannot be analyzed across grade levels.

DIBELS

DIBELS is a set of short, standardized, and individually administered assessments designed to monitor the development of early literacy and reading skills in students from kindergarten through Grade 8. Subtests vary by grade level but assess key components of early literacy such as phonemic awareness, alphabetic principle, oral reading fluency, and comprehension. Students receive subtest scores and a composite score reported as scale scores, benchmark status, and percentile rank. Percentile ranks were converted to NCEs for analyzing across grade levels in this report.

Renaissance Star

The Renaissance Star reading assessment measures students' overall reading comprehension and foundational reading skills. Students receive a set of scores including a scale score and percentile rank. Percentile ranks were converted to NCEs for analyzing across grade levels in this report.

ANALYSES

Paired t-test analyses were used to assess middle- to end-of-year growth in assessment scores. Repeated measure ANOVAs were used to assess between-group differences in gains.

EVALUATION FINDINGS

1 Do ELD students receiving Express! instruction demonstrate statistically significant gains on standardized reading tests?

The full sample of ELD students receiving Express! instruction included 206 students. A total of 182 kindergarten–Grade 4 students had both middle- and end-of year DIBELS Composite scores and 93 students in Grades 2-4 had Star Reading assessment data.

Exhibit 3 presents DIBELS Composite Reading NCE and scale scores by grade level and results of paired t-test analyses. These data show significant scale score gains at all grade levels and significant gains in NCE scores for Grade 1. Results demonstrate that while Express! students in all grade levels showed significant scale score gains—similar to students nationally—the gains of Grade 1 students exceeded those of Grade 1 students nationally.

Exhibit 3. Express! Students' DIBELS Composite Reading Assessment Scores

Grade	*N*	Middle-of-Year		End-of-Year		Change		*t*	*p*	*d*
		M	*SD*	*M*	*SD*	*M*	*SD*			
DIBELS NCE Scores										
Kindergarten	42	44.02	12.88	47.34	16.05	3.31	16.93	1.27	.212	0.20
Grade 1	40	46.77	17.34	50.19	20.66	3.42	9.52	2.27	.029	0.36
Grade 2	32	47.15	20.53	46.99	18.50	-0.16	9.75	-0.09	.925	-0.02
Grade 3	42	50.42	18.95	49.04	19.39	-1.39	9.32	-0.96	.341	-0.15
Grade 4	26	47.65	17.97	45.29	16.38	-2.37	7.00	-1.72	.097	-0.34
Total	**182**	**47.17**	**17.47**	**48.00**	**18.27**	**0.83**	**11.53**	**0.97**	**.333**	**0.07**
DIBELS Scale Scores										
Kindergarten	42	273.40	26.02	427.48	32.48	154.07	34.72	28.76	.000	4.61
Grade 1	40	328.53	21.33	463.75	47.32	135.23	30.50	28.04	.000	4.43
Grade 2	32	325.72	29.17	438.97	34.08	113.25	15.13	42.35	.000	7.37
Grade 3	42	333.52	31.24	442.57	34.85	109.05	15.15	46.64	.000	7.00
Grade 4	26	329.08	30.04	432.92	27.81	103.85	12.31	43.03	.000	4.97

Notes. Red highlighted cells show change that is statistically significant, as assessed using paired t-tests. While traditional interpretations of Cohen's *d* suggest 0.20 to be a small effect, 0.50 to be a medium effect, and 0.80 to be a large effect, current education researchers recommend interpreting effect sizes based on the study's context and caution against broadly applying Cohen's interpretations to education studies: "Effects that are small by Cohen's standards are large relative to the impacts of most field-based interventions."[2]

Exhibit 4 presents Renaissance Star Reading NCE and scale scores by grade level. These data show significant scale score gains for Grades 2 and 3 (but not Grade 4) and paired t-tests revealed significant gains in NCE scores overall and for Grade 2. Results demonstrate that while Express! students in Grades 2 and 3, but not Grade 4, showed significant progress, progress was on par with (Grades 3 and 4) or greater than (Grade 2 and Grades 2-4 as a whole) that of their national peers.

[2]Kraft, Matthew. (2020). Interpreting Effect Sizes of Education Interventions. (EdWorkingPaper: 19-10). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/8pjp-2z74

Exhibit 4. Express! Students' Renaissance Star Reading Assessment Scores

Grade	N	Middle-of-Year		End-of-Year		Change		t	p	d
		M	SD	M	SD	M	SD			
Renaissance Star NCE Scores										
Grade 2	35	36.09	18.91	39.39	15.73	3.29	8.85	2.20	.035	0.37
Grade 3	38	41.32	17.97	43.11	19.05	1.79	9.02	1.22	.229	0.20
Grade 4	20	40.84	19.02	41.13	16.18	0.29	9.14	0.14	.889	0.03
Total	**93**	**39.25**	**18.52**	**41.28**	**17.15**	**2.03**	**8.95**	**2.19**	**.031**	0.23
Renaissance Star Scale Scores										
Grade 2	35	174.43	119.60	225.14	120.62	50.71	53.74	5.58	.000	0.94
Grade 3	38	305.58	149.81	353.71	170.98	48.13	84.60	3.51	.001	0.57
Grade 4	20	404.80	182.80	430.65	164.05	25.85	84.49	1.37	.187	0.31

Note. Renaissance Star was not administered to kindergarten and Grade 1 students. Red highlighted cells show change that is statistically significant, as assessed using paired t-tests. While traditional interpretations of Cohen's *d* suggest 0.20 to be a small effect, 0.50 to be a medium effect, and 0.80 to be a large effect, current education researchers recommend interpreting effect sizes based on the study's context and caution against broadly applying Cohen's interpretations to education studies.

2 Do gains on standardized reading tests of ELD students receiving Express! instruction differ by starting performance level?

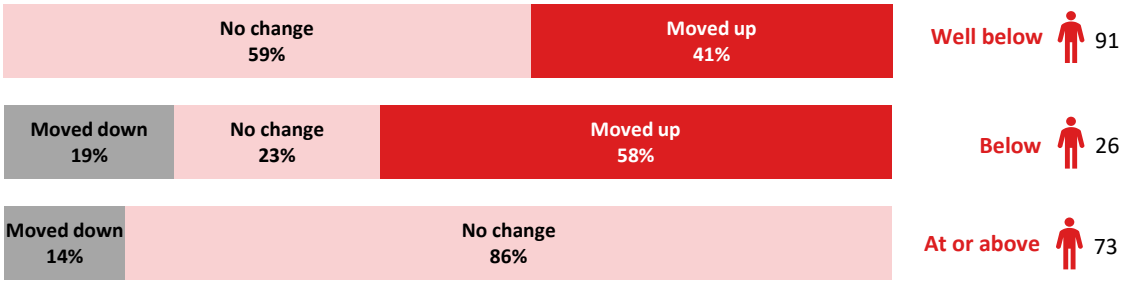
Of 190 students with DIBELS composite scores at both middle and end of year, 48% were well below benchmark, 14% were below benchmark, and 39% were at benchmark (see Exhibit 5).

Exhibit 5. DIBELS Middle-of-Year Reading Benchmark Status

Well below benchmark (Intensive Support)	Below benchmark (Strategic Support)	At or above benchmark (Core Support)
48%	14%	39%
n = 91	*n* = 26	*n* = 73

Exhibit 6 shows that a significant proportion of students who started well below or below benchmark moved toward or reached benchmark by the end of the year. A small proportion who started at benchmark at middle-of-year assessment moved down below benchmark at the end of the year.

Exhibit 6. DIBELS Change in Reading Benchmark Status of Express! Students from Middle to End of Year



The data in Exhibit 7 show that students in the lower-performing DIBELS groups—those well below and below benchmark at the middle of the year—made the greatest gains from middle- to end-of-year on both the DIBELS Composite and Star Reading assessments. Students already at benchmark showed some decrease on the DIBELS Composite Reading and no change on the Star Reading assessment. These findings suggest that Express! instruction had the strongest impact on students with the greatest need.

Exhibit 7. Express! Students' NCE Scores by DIBELS Middle-of-Year Reading Benchmark Status

Grade	*n*	Middle-of-Year		End-of-Year		Change		*F*	*p*	η^2
		M	*SD*	*M*	*SD*	*M*	*SD*			
DIBELS NCE Scores										
Well below	91	19.56	12.41	26.37	17.93	2.96	11.31	5.48	.005	0.06
Below	26	45.88	8.16	50.65	22.91	3.48	12.68			
At or above	73	72.47	13.73	67.11	19.43	-2.54	10.64			
Renaissance Star NCE Scores										
Well below	37	23.89	13.19	27.82	11.44	3.93	9.18	6.21	.015	0.07
Below	14	39.57	10.18	44.51	11.86	4.94	7.52			
At or above	38	53.79	13.58	52.61	14.93	-1.18	8.71			

Note. Renaissance Star was not administered to kindergarten and Grade 1 students. Red highlighted cells show that group differences in gains, as assessed using repeated measure ANOVA, are statistically significant. η^2 = Partial eta^2 (magnitude of the program effect). η^2 = 0.01 indicates a small effect, η^2 = 0.06 indicates a medium effect, η^2 = 0.14 indicates a large effect.

3 Do ELD students receiving Express! instruction demonstrate statistically greater growth or higher performance on standardized reading tests than an equivalent group of students not using the program?

To assess whether Express! students in Attleboro Public Schools made achievement gains that were greater than those of a demographically similar group of ELD students not receiving Express! instruction, we used a matched-comparison group design. A total of 65 Express! students (approximately one-third of the overall sample) had both middle- and end-of-year DIBELS Composite Reading scores and were able to be matched to students in the comparison group (see Exhibit 2 for Express! and comparison sample demographics). Exhibit 8 presents middle- and end-of-year mean DIBELS Composite Reading scores and mean change between time points for Express! and comparison students.

Exhibit 8. Matched Express! and Comparison Students' DIBELS Reading Assessment Scores

Grade	*n*	Middle-of-Year		End-of-Year		Change		*F*	*p*	η^2
		M	*SD*	*M*	*SD*	*M*	*SD*			
DIBELS NCE Scores										
Kindergarten										
Express!	13	44.40	11.84	48.62	16.51	4.22	18.73	1.60	.218	0.06
Comparison	13	44.53	13.56	41.13	15.17	-3.40	10.98			
Grade 1										
Express!	11	40.31	16.54	41.36	20.32	1.05	8.04	-0.46	.505	-0.02
Comparison	11	36.81	19.06	40.84	14.01	4.03	12.10			

Exhibit 8 (continued)

Exhibit 8 Continued

Grade	*n*	Middle-of-Year		End-of-Year		Change		*F*	*p*	η^2
		M	*SD*	*M*	*SD*	*M*	*SD*			
DIBELS NCE Scores										
Grade 2										
Express!	15	48.57	20.85	47.85	20.68	-0.73	8.62	0.38	.541	0.01
Comparison	15	49.32	20.21	46.77	18.19	-2.55	7.50			
Grade 3										
Express!	16	49.11	17.73	49.20	19.69	0.09	9.22	0.64	.430	0.02
Comparison	16	52.93	20.79	49.72	19.26	-3.21	13.69			
Grade 4										
Express!	10	31.65	12.16	33.31	11.56	1.66	7.05	0.10	.760	0.01
Comparison	10	33.06	19.08	33.64	16.24	0.58	8.44			
Total										
Express!	**65**	**43.87**	**17.21**	**45.00**	**18.75**	**1.13**	**11.05**	**1.58**	**.211**	**0.01**
Comparison	**65**	**44.63**	**19.67**	**43.34**	**17.32**	**-1.29**	**10.92**			
DIBELS Scale Scores										
Kindergarten										
Express!	13	272.92	24.36	429.15	30.56	156.23	34.81	1.68	.208	0.07
Comparison	13	274.38	27.69	415.38	26.32	141.00	24.20			
Grade 1										
Express!	11	320.82	17.12	444.45	45.39	123.64	29.48	0.01	.912	0.00
Comparison	11	318.36	17.25	440.82	32.09	122.45	19.21			
Grade 2										
Express!	15	327.20	30.67	439.87	37.91	112.67	15.37	0.14	.712	0.00
Comparison	15	327.07	30.06	437.87	33.60	110.80	11.86			
Grade 3										
Express!	16	332.00	29.13	443.31	36.57	111.31	15.30	0.55	.465	0.02
Comparison	16	337.50	33.12	443.94	35.30	106.44	21.48			
Grade 4										
Express!	10	301.80	22.83	412.30	26.41	110.50	15.39	0.31	.583	0.02
Comparison	10	304.10	32.55	410.40	32.69	106.30	18.06			

Notes. Group differences in gains were assessed using repeated measure ANOVAs. η^2 = Partial eta^2 (magnitude of the program effect). η^2 = 0.01 indicates a small effect. η^2 = 0.06 indicates a medium effect. η^2 = 0.14 indicates a large effect.

Exhibit 9 presents middle- and end-of-year mean Renaissance Star Reading scores and mean change between time points for Express! and comparison students. Results presented in Exhibit 8 and Exhibit 9 show no significant differences between groups for either reading assessment at any grade level. However, while not statistically significant, with the exception of Grade 4 Star scale scores, Express! students made slightly greater gains than comparison students on both reading assessments.

Exhibit 9. Matched Express! and Comparison Students' Renaissance Star Reading Assessment Scores

Grade	*n*	Middle-of-Year		End-of-Year		Change		*F*	*p*	η^2
		M	*SD*	*M*	*SD*	*M*	*SD*			
Star NCE Scores										
Grade 2										
Express!	15	33.93	18.85	38.81	15.95	4.88	9.12	1.89	.180	0.06
Comparison	15	42.40	17.82	43.30	13.78	0.90	6.54			
Grade 3										
Express!	13	42.12	18.67	43.75	19.92	1.64	7.67	0.00	.972	0.00
Comparison	16	39.81	23.96	41.36	21.66	1.54	6.87			
Grade 4										
Express!	8	31.81	15.82	30.08	12.17	-1.74	9.41	0.01	.933	0.00
Comparison	10	35.37	14.49	33.27	17.76	-2.10	8.51			
Total										
Express!	**36**	**36.41**	**18.20**	**38.65**	**17.15**	**2.24**	**8.82**	**1.00**	**.321**	**0.01**
Comparison	**41**	**39.68**	**19.52**	**40.10**	**18.15**	**0.42**	**7.15**			
Star Scale Scores										
Grade 2										
Express!	15	157.67	112.65	216.47	128.55	58.80	53.83	0.89	.355	0.03
Comparison	15	209.73	133.13	250.13	115.24	40.40	53.24			
Grade 3										
Express!	13	311.15	158.73	354.08	179.22	42.92	65.41	0.30	.591	0.01
Comparison	16	299.19	196.11	330.44	194.72	31.25	50.32			
Grade 4										
Express!	8	320.50	132.13	320.75	121.46	0.25	82.55	-0.03	.873	-0.00
Comparison	10	343.50	141.07	350.70	184.87	7.20	96.08			

Notes. Renaissance Star was not administered to kindergarten and Grade 1 students. Group differences in gains were assessed using repeated measure ANOVAs. η^2 = Partial eta^2 (magnitude of the program effect). η^2 = 0.01 indicates a small effect. η^2 = 0.06 indicates a medium effect. η^2 = 0.14 indicates a large effect.

4 Do reading achievement gains vary based on student demographic characteristics?

To assess whether Express! differentially impacted students of different demographic subgroups, we assessed reading assessment gains by groups. Exhibit 10 presents middle- and end-of-year mean scores and mean change between time points and group difference statistics. The data show no significant differences between gender, Hispanic ethnicity, special education status, or socioeconomic status subgroups for either reading assessment.

Exhibit 10. Express! Students' Reading Assessment NCE Scores by Demographic Subgroup

Grade	*n*	Middle-of-Year		End-of-Year		Change		*F*	*p*	η^2
		M	*SD*	*M*	*SD*	*M*	*SD*			
DIBELS NCE Scores										
Gender										
Male	92	47.13	18.54	48.62	20.47	1.49	11.40	0.61	.436	0.00
Female	90	47.22	16.41	47.37	15.79	0.15	11.68			
Ethnicity										
Non-Hispanic	96	48.71	17.19	49.39	17.84	0.68	11.24	0.03	.857	0.00
Hispanic	86	45.46	17.73	46.45	18.71	0.99	11.90			
Special Education										
No	144	49.55	17.31	50.19	17.83	0.65	11.55	0.17	.679	0.00
Yes	38	38.18	15.17	39.70	17.72	1.52	11.56			
Free- or Reduced-Price Meals										
No	41	51.50	17.12	51.61	15.85	0.11	9.83	0.21	.651	0.00
Yes	141	45.91	17.43	46.95	18.83	1.04	12.00			
Star NCE Scores										
Gender										
Male	50	40.71	19.90	42.34	18.51	1.63	8.55	0.22	.640	0.00
Female	43	37.55	16.84	40.05	15.56	2.50	9.49			
Ethnicity										
Non-Hispanic	46	40.87	18.76	42.88	17.33	2.01	8.44	0.00	.981	0.00
Hispanic	47	37.66	18.34	39.72	17.02	2.06	9.52			
Special Education										
No	71	42.85	18.42	44.39	17.54	1.55	9.08	0.89	.349	0.01
Yes	22	27.65	13.65	31.25	11.21	3.60	8.54			
Free- or Reduced-Price Meals										
No	25	49.06	14.93	49.12	14.64	0.07	8.27	1.66	.201	0.02
Yes	68	35.64	18.49	38.40	17.20	2.76	9.14			

Notes. Renaissance Star was not administered to kindergarten and Grade 1 students. Group differences in gains were assessed using repeated measure ANOVAs.

SUMMARY AND CONCLUSIONS

This evaluation was conducted with a sample of students in Attleboro Public Schools in Attleboro, Massachusetts, from three of the district's five schools that piloted Benchmark Education Company's Express! instructional program in spring 2025 for students enrolled in ELD programs. A total of 206 students in kindergarten through Grade 4 received Express! instruction. Students in two of the district's other schools were eligible to serve as a comparison group. Due to demographic dissimilarity between groups, a reduced matched-comparison group comprising approximately one-third of the Express! sample, was established for this evaluation.

Renaissance Star and DIBELS assessments were used to assess student growth in reading and language skills between the middle and the end of the year. Key findings are highlighted below.

Do ELD students receiving Express! instruction demonstrate statistically significant gains on standardized reading tests?

Express! students in all grade levels showed significant scale score gains on the DIBELS Composite Reading assessment, indicating that students in all grades made significant reading achievement progress. For the overall sample of students, NCE growth on DIBELS was nonsignificant, indicating that Attleboro Express! students progressed at a rate that was similar to K-4 students nationally. For Grade 1 students specifically, however, NCE growth on DIBELS was significant, indicating that Attleboro Grade 1 students made greater progress that Grade 1 students nationally.

Star Reading scale score gains were significant for Grades 2 and 3 but not for Grade 4, indicating that while Grade 2 and Grade 3 Attleboro Express! students made significant reading progress on the Star assessment, Grade 4 students did not; however, NCE growth for Grade 4 was also nonsignificant, meaning that although these students did not demonstrate significant scale score gains, their progress was on par with Grade 4 students nationally. For the overall sample of students, NCE growth on Star was significant, signifying that Attleboro Express! students as a whole made greater progress than students in Grades 2-4 nationally. Additionally, NCE growth on Star was significant for Grade 2 students specifically, indicating that Attleboro Grade 2 students made greater progress than their national Grade 2 peers.

Do gains on standardized reading tests of ELD students receiving Express! instruction differ by starting performance level?

On the DIBELS assessment, 41% of students who started well below and 58% who started below benchmark—a significant proportion—moved toward or reached benchmark by the end of the year; in contrast, for those who started at benchmark, most (86%) remained at benchmark at the end of the year. While overall DIBELS NCE growth between middle- and end-of year wasn't statistically significant, differences by starting proficiency level were. Specifically, students performing below or well below benchmark at start showed significantly greater growth on NCE scores than students who started at benchmark. This finding was similarly true for the Star Assessment. These findings suggest that the Express! instruction had the strongest impact on students with the greatest need and that the higher need students made greater progress than (a) their higher achieving Attleboro peers and (b) same-grade national peers.

Do ELD students receiving Express! instruction demonstrate statistically greater growth or higher performance on standardized reading tests than an equivalent group of students not using the program?

Students in the Express! subsample of students performed similarly to a matched subsample of students not participating in Express! instruction on both the DIBELS and Star reading assessments. However, while for most grade levels, Express! students performed slightly better than comparison students, reduced sample sizes limited statistical power to detect differences.

Do reading achievement gains vary based on student demographic characteristics?

Reading achievement gains on DIBELS and Renaissance Star assessments did not differ by gender, Hispanic ethnicity, special education status, or socioeconomic status.

RECOMMENDATIONS

This evaluation offers promising findings for Benchmark Education Company's Express! instructional program for students enrolled in ELD programs, and results are particularly promising as they relate to high need students (i.e., those performing below or well below benchmark). The study was strengthened by its use of nationally normed assessments and a reasonable sample size of Express! participants.

Although it was the intent of this evaluation to compare Express! outcomes to a similar group of ELD students receiving their business-as-usual instruction, the demographic characteristics between groups were ultimately too dissimilar to draw meaningfully comparisons. Given the small comparison student pool, a one-to-one matched comparison sample yielded a subsample that was only one-third of the original Express! sample. A more rigorous evaluation design—a randomized control trial, if feasible, or a quasi-experimental design in which comparison students can be matched from within the same school or a larger pool of potential comparison students, more generally—would allow for more meaningful assessment of program impacts.

Based on the findings of this single district evaluation, further exploration on the benefits of Express! for students performing below benchmark is also warranted.

APPENDIX A. EVALUATION DESIGN RATIONALE

A quasi-experimental design was the original approach for the evaluation, with Express! being implemented to K-4 students in three of Attleboro's five elementary schools and a comparison group including K-4 ELD students in the district's remaining two elementary schools. However, post-implementation baseline equivalence analyses revealed substantial and statistically significant differences between the Express! and comparison groups on several key demographic indicators and on the baseline assessment—with differences that differed by grade level (see Exhibit A1).

Exhibit A1. Intended Sample: Baseline Equivalence Between Express! and Comparison Group

	Kindergarten		Grade 1		Grade 2		Grade 3		Grade 4	
Variable	**Express!**	**Comp**	**Express!**	**Comp**	**Express!**	**Comp**	**Express!**	**Comp**	**Express!**	**Comp**
Sample size (*n*)	47	25	44	25	37	34	47	17	31	25
Percent Female	53%	44%	50%	32%	41%	47%	53%	53%	52%	48%
Percent AIAN	21%	24%	**7%**	**36%**	19%	15%	6%	12%	6%	12%
Percent Asian	13%	20%	20%	28%	16%	6%	32%	18%	23%	12%
Percent Black	17%	8%	**32%**	**12%**	27%	21%	23%	12%	29%	20%
Percent Hispanic	53%	68%	**30%**	**76%**	46%	59%	0.53	0.71	52%	56%
Percent White	23%	24%	**20%**	**4%**	14%	21%	9%	12%	6%	16%
Percent FRM	79%	92%	84%	92%	78%	74%	**74%**	**94%**	68%	88%
Percent SPED	15%	24%	25%	12%	**32%**	**9%**	15%	18%	19%	32%
Baseline DIBELS NCE	44.02	45.83	**46.77**	**36.10**	47.15	52.18	50.79	47.66	**47.65**	**34.88**

Notes. Total Express! *n* = 206; Total comparison *n* = 142. **Red text** indicates a significant difference between the Express! and comparison group, as assessed using independent t-tests. Comp = comparison. AIAN = American Indian and Alaska Native. FRM = eligible for free- or reduced-price meals. SPED = special education.

To address this challenge, the evaluation shifted to a two-step design:

1. Pre-post design in which all Express! students were assessed on their pre- and post-assessment score gains
2. Matched-comparison design with a subsample of Express! and comparison students who were one-to-one matched on key demographic characteristics and baseline assessment scores using propensity score next-neighbor matching.

The purpose of using propensity score matching was to identify a sample of non-Express! students that most closely resembles the non-Express! students to reduce potential bias and allow for a more credible estimation of the treatment effect. A propensity score was generated based on the following covariates:

- baseline DIBELS normal curve equivalent (NCE) score,
- gender,
- race (White/non-White),
- ethnicity (Hispanic/non-Hispanic)
- special education status, and
- eligibility for free- or reduced-price meals.

Within each grade level, each ELD comparison student was matched to an Express! student based on the closest matched propensity score. From this process, we were able to create a subsample of 65 K-4 students who participated in Express! in spring 2025 and 65 matched comparison students drawn from the district's two other elementary schools. The groups were statistically equivalent on demographic characteristics and baseline assessment scores (see Exhibit A2).

Exhibit A2. Propensity Score Matched Sample: Baseline Equivalence Between Express! and Comparison Group

	Kindergarten		Grade 1		Grade 2		Grade 3		Grade 4	
Variable	**Express!**	**Comp**	**Express!**	**Comp**	**Express!**	**Comp**	**Express!**	**Comp**	**Express!**	**Comp**
Sample size (*n*)	13	13	11	11	15	15	16	16	10	10
Percent Female	54%	38%	27%	27%	33%	47%	56%	56%	40%	30%
Percent AIAN	8%	8%	18%	18%	7%	7%	25%	19%	10%	10%
Percent Asian	15%	31%	9%	36%	33%	7%	0%	13%	0%	0%
Percent Black	15%	8%	27%	27%	13%	27%	25%	19%	30%	30%
Percent Hispanic	46%	54%	64%	55%	67%	47%	63%	56%	70%	50%
Percent White	38%	38%	0%	9%	20%	27%	0.13	0.13	0%	20%
Percent FRM	100%	100%	100%	100%	73%	80%	75%	94%	80%	90%
Percent SPED	8%	38%	9%	9%	40%	13%	6%	6%	20%	30%
Baseline DIBELS NCE	44.40	44.53	40.31	36.81	48.57	49.32	49.11	52.93	31.65	33.06

Notes. PSM Express! *n* = 65; PSM comparison *n* = 65. Independent t-test analyses revealed no statistically significant differences between the Express! and comparison subsamples on demographic characteristics or baseline DIBELS scores. Comp = comparison. AIAN = American Indian and Alaska Native. FRM = eligible for free- or reduced-price meals. SPED = special education.