Impact Evaluation of the Equity Leader Accelerator Program

Technical Appendices

Martyna Citkowicz I Sarah Peko-Spicer I Jill Bowdon I Melissa Brown-Sims I Sarah Mae Olivar I Sonica Dhillon

September 2025

American Institutes for Research® | AIR.ORG

AIR®
Advancing Evidence.
Improving Lives.

Contents

Exhibits

Appendix A. Student Achievement Study

This appendix describes the study on the impact of the Center for Leadership and Educational Equity (CLEE) Equity Leader Accelerator Program (ELAP) on student achievement. We first describe our methodological approach, then the analytic sample, and, finally, the results.

Methodological Approach

In this section, we first describe the eligible samples and outcomes and then the analytic approach.

Eligible Samples and Outcomes

AIR measured student achievement using the 2023–24 Massachusetts Comprehensive Assessment System (MCAS) English language arts (ELA) and math test scores for students in Grades 3–8 and Grade 10.

To prepare the data for analysis, AIR used the ELAP School District Crosswalk to identify ELAP-participating schools. We removed 13 schools, including those that did not report school-level data, schools that serve only K–2 students, and schools that serve only 9–12 students.[1] Within the 25 remaining eligible ELAP schools, we removed 3,809 students in the ELA sample and 3,769 students in the math sample who were missing outcome and baseline achievement scores (i.e., a complete-cases approach). Exhibit A1 presents our identification process for each outcome in the administrative data, showing how the sample sizes changed as we implemented these criteria.[2]

[1] MCAS student achievement data are not available for K–2 schools. Students included in the sample had to have at least an ELA or math score from the analytic school year (2023–24) and an ELA or math score from the prior school year (2022–23). Only students in Grades 4–8 met this criterion because there is no prior-year MCAS score for students in Grades 3 and 10.

[2] The sample sizes in Exhibit A1 may differ from the analytic samples presented later in the appendix, which include schools and students for whom we were able to identify matches. The matching process is described in the next section.

Exhibit A1. Number of Eligible ELAP Schools and Students Included in the Achievement Sample, by Subject

Criteria	ELA	Math
Schools		
Number of ELAP schools in the ELAP School District Crosswalk with Early Career Principals	38	38
Number of ELAP schools identified in the data [a]	38	38
Number of ELAP schools that reported school-level data	34	34
Number of ELAP schools serving students in tested grades (Grades 3–8 and Grade 10)	30	30
Number of ELAP schools serving students in tested grades with outcomes (ELA or math score) and baseline (prior-year ELA or math score) data	25	25
Students		
Number of students in eligible ELAP schools	11,087	11,087
Number of students in eligible ELAP schools in tested grades (Grades 3–8 and Grade 10) with outcomes (ELA or math score) and baseline (prior-year ELA or math score) data	7,278	7,318

Note. ELA = English language arts.

[a] Not identified in the data means that an ELAP school was either not found in the school-level data or the student-level data.

Analytic Approach

To examine the impact of ELAP on student achievement, we followed a two-stage approach. First, we used propensity score matching to identify comparison schools in the administrative data similar to ELAP schools in terms of: (a) school level (elementary, middle, or high), (b) school locale (i.e., rural or suburban), (c) school size (based on total student enrollment), (d) MCAS proficiency rates in ELA and math in the prior school year, and (e) student composition.[3] Comparison schools followed a "business-as-usual" approach,

The goal of matching is to construct a comparison group that is the same at baseline as the intervention group (ELAP). Matching allows for an apples-to-apples comparison and enables the inference that any difference in outcomes is due to the presence of ELAP as opposed to preexisting baseline differences.

[3] The gold standard impact study design involves random assignment of schools to one of two conditions: receiving ELAP (intervention group) or not (comparison group). Because random assignment was not feasible for this study, we use a quasi-experimental approach called propensity score matching to construct a comparison group based on baseline data, including background information measured at the student and school levels. We can then contrast the outcomes of leaders and

maintaining the standard resources and conditions that would typically be available to the group targeted for the intervention.[4] We forced matches within school level (elementary, middle, or high), and then we created two analytic samples: one for ELA achievement and one for math achievement. Second, focusing solely on the schools (and the students in those schools) identified in the matching process, we estimated differences in standardized ELA and math scores between the students in ELAP schools and the students in comparison schools.

Identifying Comparison Schools

We used propensity score matching to select three comparison schools to each participating ELAP school. To implement propensity score matching, we ran the following logistic regression model:

$$logit\{P(Treatment_s)\} = \beta_0 + \boldsymbol{W_s}$$

where $Treatment_s$ is an indicator for whether school *s* was a treatment school (coded as 1 for ELAP schools and 0 for comparison schools); $P(Treatment_s)$ is the propensity of being a treatment school; and $\boldsymbol{W_s}$ is a vector of school-level characteristics (i.e., school locale, total student enrollment, prior-year ELA and math proficiency rates, and percentages of students by race/ethnicity, by gender, English learners, students with disabilities, and students from low-income households). When establishing the non-ELAP comparison group, we forced matches within school level so that participating ELAP elementary schools were matched only to other nonparticipating elementary schools; the same method was used to match middle schools with middle schools and high schools with high schools.

After matching, we created two separate samples: (a) matched sample of students who have ELA scores, and (b) matched sample of students who have math scores. We assessed baseline equivalence for both samples to examine whether each sample of students in comparison schools was indeed similar to the corresponding sample of students in ELAP schools prior to conducting the analyses.

students from the ELAP intervention group and the constructed comparison group. This contrast allows us to answer the question, "What is the impact of ELAP relative to the business-as-usual condition?"

[4] Eligible comparison schools had to meet the following three criteria: (a) the school was categorized as a regular school by the National Center for Education Statistics (i.e., not a special education, vocational, or other/alternative school), (b) the school served students in Grades 3–8 and/or 10, and (c) school data had been reported for total enrollment, student demographics, and MCAS scores in 2022–23.

Exhibits A2 presents the baseline equivalence results at the school and student levels for each adequately matched sample.[5] The values represent the standardized mean difference (SMD) between students in ELAP schools and students in matched comparison schools on their demographics, school characteristics, and prior-year achievement from the state assessments. All SMDs for prior-year achievement in the tested subject were below 0.25, the required cutoff for achieving baseline equivalence.[6] *This means that any difference in outcomes is due to the presence of ELAP, not preexisting baseline differences.*

The standardized mean difference is the average difference between treatment and comparison groups divided by the standard deviation. It tells us how different the two groups are on a specific measure.

Exhibit A2. Baseline Equivalence in Student and School Characteristics and Prior Achievement for the Achievement Outcomes Samples, by Subject

Characteristic	ELA (school *N* = 99; student *N* = 25,564)	Math (school *N* = 99; student *N* = 25,663)
School characteristics		
Rural school	0.05	0.05
Average student enrollment	0.30	0.30
Percentage of students who are White and non-Hispanic	−0.04	−0.04
Percentage of female students	−0.40	−0.40
Percentage of English learners	−0.12	−0.12
Percentage of students with disabilities	−0.09	−0.09
Percentage of students from low-income households	−0.05	−0.05
Average prior-year ELA proficiency rate	0.05	0.05
Average prior-year math proficiency rate	0.04	0.04
Student characteristics		
Students who are White and non-Hispanic	< 0.01	0.01
Female	−0.02	−0.02

[5] Although matching took place at the school level, we also assessed whether the matches were adequate at the student level (i.e., on student-level variables, including demographics and prior-year standardized achievement scores) for the student achievement analyses.

[6] Per What Works Clearinghouse (WWC) standards, baseline equivalence is achieved with an SMD of 0.25 or below; here, between students in ELAP schools and students in comparison schools on the variables included in the matching analysis, particularly those variables related to students' prior-year achievement. In addition, baseline characteristics for which the SMD is between 0.05 and 0.25 must be included in the analytic model as covariates to statistically adjust for any remaining baseline differences (WWC, 2022). We include all characteristics and prior-year achievement scores in our analytic models.

Characteristic	ELA (school *N* = 99; student *N* = 25,564)	Math (school *N* = 99; student *N* = 25,663)
English learners	0.06	0.06
Students with disabilities	−0.01	−0.01
Students from low-income households	0.06	0.06
Average prior-year ELA standardized scale score	−0.02	N/A
Average prior-year math standardized scale score	N/A	−0.03

Note. N/A = not applicable. The values represent the SMD between students in ELAP schools and students in matched comparison schools.

Analyses of Student Achievement Outcomes

To estimate differences in ELA and math standardized achievement scores between the students in ELAP schools and the students in matched comparison schools, we estimated the following multilevel model separately for each outcome:

$$Y_{is} = \beta_0 + \beta_1 Treatment_s + \boldsymbol{X_{is}} + \boldsymbol{W_s} + \boldsymbol{SchoolLevel_s} + v_s + e_{is}$$

where Y_{is} is the outcome of interest (i.e., standardized ELA or math achievement score)[7] for student *i* in school *s*; $Treatment_s$ is an indicator for whether a student was in a treatment school (coded as 1 for attending an ELAP school and 0 for attending a comparison school); $\boldsymbol{X_{is}}$ is a vector of student characteristics (i.e., prior-year ELA or math test score, race, gender, English learner status, disability status, and low-income household status); and $\boldsymbol{W_s}$ is a vector of school-level characteristics (i.e., school locale, total student enrollment, prior-year ELA and math proficiency rates, and percentages of students by race/ethnicity, by gender, English learners, students with disabilities, and students from low-income households). School level ($\boldsymbol{SchoolLevel_s}$) fixed effects are included to ensure that students only within the same school level are compared with one another. Random effects are included to account for the residual effects of schools (v_s) and students (e_{is}). The main parameter of interest is β_1, which represents the estimated mean difference in the outcome measure (i.e., achievement scores) between students in ELAP schools and students in matched comparison schools.

We also conducted exploratory analyses to examine whether program impact on ELA and math achievement differed by school level and student high-need status by including an interaction between the treatment and indicator of group membership and running the analytic model separately for each interaction. For example, for school level, we included the interaction term

[7] To ensure that scores from different assessments or years were on the same scale, we standardized achievement scores by district, grade, subject, and year.

$Treatment_s * \boldsymbol{SchoolLevel_s}$ in the model. Because matching was not conducted for each individual group, these group analyses should be interpreted with caution because group membership may be associated with, or a proxy for, other factors producing differences.

To examine program performance on standardized ELA and math achievement scores for "focal" students, we conducted a second set of exploratory analyses focused solely on ELAP schools. Specifically, within ELAP schools, we examined whether program impact on ELA and math achievement differed between focal students and non-focal students. In each ELAP school, Early Career Principals identified "focal" students who received targeted help and support. These students often included English learners, students with disabilities, those who identify as African American/Black or Hispanic, and students in specific grades. These analyses do not include a comparison group that was not exposed to ELAP, and matching was not conducted to identify the non-focal students; thus, these analyses are considered exploratory and should be interpreted with caution.

Finally, to examine the impact of ELAP on all participating schools, we conducted exploratory analyses on the broader sample of schools using school-level data. The student-level data did not allow us to include high school grades in the achievement analyses due to missing prior-year scores. By using school-level data, we were able to include schools across all grade levels.

Transforming Effects

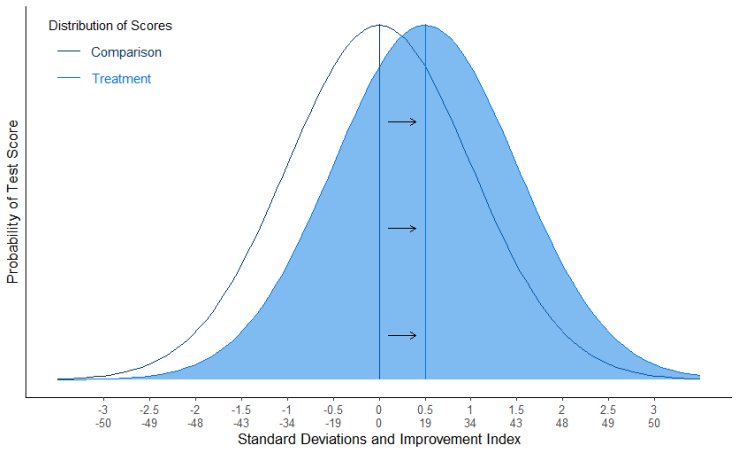
To facilitate interpretation, all treatment estimates for achievement outcomes were transformed into the WWC improvement index metric (Hedges & Olkin, 2016; WWC, 2022) using the following formula:

$$improvement\ index = 100\Phi(estimate) - 50$$

where $\Phi(estimate)$ denotes the standard normal cumulative distribution function evaluated at the treatment effect *estimate*. The improvement index represents the difference between the percentile rank corresponding to the treatment group mean and the percentile rank corresponding to the comparison group mean (that is, the 50th percentile) based on the comparison group distribution.

WHAT IS THE WHAT WORKS CLEARINGHOUSE (WWC) IMPROVEMENT INDEX?

The WWC improvement index represents the difference between the percentile rank corresponding to the treatment group mean and the percentile rank corresponding to the comparison group mean (at the 50th percentile) based on the comparison group distribution. In the figure below, the treatment group distribution (light blue line with shading) is shifted to the right of the comparison group distribution (dark blue line with no shading). An effect of 0.5 *SD* indicates an increase of a student's score from the 50th percentile to the 69th percentile, resulting in an improvement index of +19. This value indicates that the average student from the comparison group would have scored 19 percentile points higher in achievement if they had been taught by a teacher from the treatment group.



Analytic Samples

Exhibit A3 shows the number of schools and students comprising the analytic samples for the study on student achievement outcomes, including a total of 99 schools with 25,564 students in the ELA sample and 25,663 students in the math sample. The ELA and math student samples allow us to detect a minimum detectable effect size of 0.03.[8] This indicates that our study is well-powered and sufficient to detect effects that are at or below the magnitude of those typically found in the literature, including those considered small but educationally meaningful

[8] For student achievement, the two-level power analysis with treatment at the cluster (or school) level includes the following based on the analytic sample and model: α = .05; a two-tailed test; power = .80; percentage of Level 2 units assigned to treatment (ELAP) = 28%; intraclass correlation coefficient (ICC) = .02; R^2 Level 1 (student) = .66; R^2 Level 2 (school) = .95; 258 students per school, on average.

by benchmarks proposed by Kraft (2020). Moreover, a study on the McREL Balanced Leadership program found effect sizes on achievement that ranged from –0.02 to 0.04, although these were not statistically significant (Jacob et al., 2015).

Exhibit A3. Number of Schools and Students in the Achievement Outcomes Samples

Sample	ELAP schools	Students in ELAP schools	Comparison schools	Students in comparison schools
ELA sample	25	7,278	74	18,286
Math sample	25	7,318	74	18,345

Exhibits A4 and A5 present the demographic characteristics for the ELA and math student achievement outcomes samples.

ELA. The ELA student sample consisted of two study groups, with 7,278 students in 25 ELAP schools and 18,286 students in 74 comparison schools. Students in both study groups were mostly White and non-Hispanic (65% and 65%), and about half were female (48% and 49%). About one-third of the students were from low-income households (36% and 33%), one-fifth had disabilities (20% and 21%), and few were English learners (9% and 7%). In addition, most students were in middle school grades (64% and 63%) and few schools were in rural settings (8% and 7%).

Math. The math student sample consisted of two study groups, with 7,318 students in 25 ELAP schools and 18,345 students in 74 comparison schools. Similar to the ELA students, most math students were White and non-Hispanic (65% and 64%), about half were female (48% and 49%), one-third were from low-income households (36% and 33%), one-fifth had disabilities (20% and 21%), and few were English learners (9% and 8%). In addition, most students were in middle school grades (63% and 63%) and few schools were in rural settings (8% and 7%).

Exhibit A4. Student and School Characteristics in the English Language Arts Achievement Sample, by Study Group

Characteristics	Students in ELAP schools (student *N* = 7,278; school *N* = 25)	Students in comparison schools (student *N* = 18,286; school *N* = 74)
Student characteristics		
Students who are White and non-Hispanic	65%	65%
Female	48%	49%
English learners	9%	7%
Students with disabilities	20%	21%
Students from low-income households	36%	33%

Characteristics	Students in ELAP schools (student *N* = 7,278; school *N* = 25)	Students in comparison schools (student *N* = 18,286; school *N* = 74)
School grade level		
Elementary (Grades 4–5)	36%	37%
Middle (Grades 6–8)	64%	63%
School characteristics		
Rural school	8%	7%
Average student enrollment	522 students	460 students
Percentage of students who are White and non-Hispanic	63%	64%
Percentage of female students	48%	49%
Percentage of English learners	9%	10%
Percentage of students with disabilities	18%	19%
Percentage of students from low-income households	33%	35%
Average prior-year ELA proficiency score	497.95	497.45
Average prior-year math proficiency score	498.01	497.59

Exhibit A5. Student and School Characteristics in the Math Achievement Sample, by Study Group

Characteristics	Students in ELAP schools (student *N* = 7,318; school *N* = 25)	Students in comparison schools (student *N* = 18,345; school *N* = 74)
Student characteristics		
Students who are White and non-Hispanic	65%	64%
Female	48%	49%
English learners	9%	8%
Students with disabilities	20%	21%
Students from low-income households	36%	33%
School grade level		
Elementary (Grades 4–5)	37%	37%
Middle (Grades 6–8)	63%	63%
School characteristics		
Rural school	8%	7%
Average student enrollment	522 students	460 students
Percentage of students who are White and non-Hispanic	63%	64%
Percentage of female students	48%	49%

Characteristics	Students in ELAP schools (student *N* = 7,318; school *N* = 25)	Students in comparison schools (student *N* = 18,345; school *N* = 74)
Percentage of English learners	9%	10%
Percentage of students with disabilities	18%	19%
Percentage of students from low-income households	33%	35%
Average prior-year ELA proficiency score	497.95	497.45
Average prior-year math proficiency score	498.01	497.59

Results

Exhibits A6 and A7 present the overall and differential impact estimates of ELAP on ELA and math achievement. The estimates may be interpreted as follows.

A value of −0.02 for overall ELA achievement, for example, indicates that students in ELAP schools, on average, had ELA scores that were 0.02 standard deviations (*SD*) lower than students in comparison schools. Translated into the WWC improvement index, the average student from the non-ELAP comparison school would have scored −0.89 percentile points lower in ELA achievement if they had attended an ELAP school. However, this estimate is not statistically significant and suggests that the average difference in student ELA achievement scores is not conclusively different between the two groups.

For analyses of ELAP schools only, a value of 0.05 for ELA achievement indicates that focal students in ELAP schools, on average, had ELA scores that were 0.05 *SD* higher than non-focal students in ELAP schools. Translated into the WWC improvement index, the average non-focal student in an ELAP school would have scored +2.04 percentile points higher in ELA achievement if they had been a focal student in an ELAP school.

Exhibit A6. Impact Estimates of ELAP on English Language Arts

Group	*N*	Estimate	*SE*	*p* value	Improvement index
Overall	25,564	−0.02	0.02	0.34	−0.89
School grade level					
Elementary (Grades 4–5)	13,968	0.01	0.03	0.69	+0.47
Middle (Grades 6–8)	11,596	−0.08	0.04	0.04	−3.14
Student high-need status [a]					
High-need	11,824	−0.02	0.03	0.54	−0.68
Not high-need	13,740	−0.03	0.03	0.33	−1.05

Group	*N*	Estimate	*SE*	*p* value	Improvement index
Focal group status (ELAP schools only) [b]					
Focal students	6,309	0.05	0.03	0.05	+2.04

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.
[a] MA DESE (2024) defines "high-need" students as those from low-income households, English learners (current or former), or students with disabilities. [b] This analysis focused solely on ELAP schools. In each ELAP school, Early Career Principals identified "focal" students who received targeted help and support. These students often included English learners, students with disabilities, those who identify as African American/Black or Hispanic, and students in specific grades.

Exhibit A7. Impact Estimates of ELAP on Math

Group	*N*	Estimate	*SE*	*p* value	Improvement index
Overall	25,663	0.02	0.02	0.34	+0.88
School grade level					
Elementary (Grades 4–5)	14,032	0.06	0.03	0.03	+2.52
Middle (Grades 6–8)	11,631	−0.05	0.04	0.18	−1.92
Student high-need status [a]					
High-need	11,912	0.03	0.03	0.23	+1.28
Not high-need	13,751	0.01	0.03	0.62	+0.52
Focal group status (ELAP schools only) [b]					
Focal students	6,008	−0.05	0.02	0.02	−2.14

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.
[a] MA DESE (2024) defines "high-need" students as those from low-income households, English learners (current of former), or students with disabilities. [b] This analysis focused solely on ELAP schools. In each ELAP school, Early Career Principals identified "focal" students who received targeted help and support. These students often included English learners, students with disabilities, those who identify as African American/Black or Hispanic, and students in specific grades.

Exhibits A8 and A9 present the overall and differential impact estimates of ELAP on ELA and math achievement using school-level data from the broader sample of schools, which allows for the inclusion of high schools. The estimates may be interpreted similarly to those above.

A value of −0.02 for overall ELA achievement indicates that ELAP schools, on average, had ELA student scores that were 0.02 *SD* lower than comparison schools. Translated into the WWC improvement index, the average non-ELAP comparison school would have scored −0.84

percentile points lower in ELA student achievement if they had been an ELAP school. However, this estimate is not statistically significant.

These analyses show that the pattern of school-level results using the broader sample is similar to the student-level results for ELA achievement and differs minimally for math achievement. For ELA achievement, the overall estimates and the group estimates are very similar across the student- and school-level analyses, with no change in the direction of the effects; none of the estimates for the school-level analyses are statistically significant. For math achievement, the addition of high schools changes the overall estimate from 0.02 *SD* for the student-level analysis to −0.03 *SD* for the school-level analysis; however, neither was statistically significant and both are very close to zero. The math estimates for the high-need analyses also changed direction from positive in the student-level analyses (high-need = 0.03 *SD*; not high-need = 0.01 *SD*) to negative in the school-level analyses (high-need = −0.05 *SD*; not high-need = −0.01 *SD*); however, neither set of analyses was statistically significant. The direction of the math effects did not change for the various school levels.

Exhibit A8. Impact Estimates of ELAP on English Language Arts—Broader School-Level Sample

Group	*N*	Estimate	*SE*	*p* value	Improvement index
Overall	118	−0.02	0.08	0.79	−0.84
School level					
Elementary	62	0.01	0.11	0.89	+0.58
Middle	32	−0.16	0.16	0.29	−6.54
High	24	0.07	0.18	0.69	+2.84
School high-need status [a]					
High-need	43	< 0.01	0.10	1.00	< 0.01
Not high-need	75	−0.07	0.13	0.62	−2.62

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.

[a] Schools with 50% or more high-need students were classified as "high need."

Exhibit A9. Impact Estimates of ELAP on Math—Broader School-Level Sample

Group	*N*	Estimate	*SE*	*p* value	Improvement index
Overall	118	−0.03	0.08	0.74	−1.11
School level					
Elementary	62	0.09	0.11	0.40	+3.75
Middle	32	−0.17	0.16	0.28	−6.79
High	24	−0.16	0.18	0.37	−6.51
School high-need status [a]					
High-need	43	−0.05	0.10	0.58	−2.16
Not high-need	75	−0.01	0.13	0.93	−0.42

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.

[a] Schools with 50% or more high-need students were classified as "high need."

Appendix B. School Climate Study

This appendix describes the study on the impact of the Center for Leadership and Educational Equity (CLEE) Equity Leader Accelerator Program (ELAP) on school climate. We first describe our methodological approach, then the analytic sample, and, finally, the results.

Methodological Approach

In this section, we first describe the eligible sample and outcomes and then the analytic approach.

Eligible Sample and Outcomes

To measure school climate, AIR used the spring 2023–24 Views of Climate and Learning (VOCAL) survey data for students in Grades 4, 5, 8, and 10 from MA DESE. The VOCAL survey measures students' perception of school climate along three dimensions: Engagement, Safety, and Environment (MA DESE, 2024). The topics measured within each dimension are presented in Exhibit B1.

Exhibit B1. Topics Measured Within Each of the Three VOCAL Survey Dimensions

Engagement	Safety	Environment
Cultural Competence—The extent to which students feel adults/students value diversity, manage dynamics of differences, and avoid stereotypes.	**Emotional Safety**—The extent to which students feel a bond to the school, and the extent to which adults/students support the emotional needs of students.	**Instructional Environment**—The extent to which students feel the instructional environment is collaborative, relevant, challenging, and supportive of learning.
Participation—The extent to which students feel engaged intellectually, emotionally, and behaviorally in the classroom, and the extent to which students or their parents are engaged in school life.	**Physical Safety**—The extent to which students feel physically safe within the school environment.	**Mental Health**—The extent to which students have access to support systems that effectively support their social, emotional, and mental health well-being.
Relationships—The extent to which students feel there is a social connection and respect between staff/teachers and students, and between students and their peers.	**Bullying**—The extent to which students report different types of bullying behaviors occurring in the school and the extent to which school/staff/students try to counteract bullying.	**Discipline Environment**—The extent to which discipline is fair, applied consistently and evenly, and a shared responsibility among staff, teachers, and students.

To prepare the data for analysis, AIR used the ELAP School District Crosswalk to identify ELAP-participating schools. We removed 10 schools, including those that did not report school-level

data and schools that serve only K–2 students.[9] Exhibit B2 presents our identification process in the administrative data, showing how the sample size changed as we implemented these criteria.[10]

Exhibit B2. Number of Eligible ELAP Schools Included in the School Climate Sample

Criteria	School climate
Number of ELAP schools in the ELAP School District Crosswalk with Early Career Principals	38
Number of ELAP schools identified in the data [a]	38
Number of ELAP schools that reported school-level data	34
Number of Grade 3 and up schools	30
Number of schools with VOCAL survey data	28

[a] Not identified in the data means that an ELAP school was not found in the school-level data.

Analytic Approach

To examine the impact of ELAP on school climate, we followed a two-stage approach similar to what was described in Appendix A for the student achievement study. First, we used propensity score matching to identify comparison schools[11] in the administrative data similar to ELAP schools in terms of: (a) school level (elementary, middle, or high), (b) school locale (i.e., rural or suburban), (c) school size (based on total student enrollment), (d) average VOCAL survey scores in the prior school year, and (e) student composition. Second, focusing solely on the schools identified in the matching process, we estimated differences in VOCAL survey outcomes between ELAP schools and comparison schools.

Identifying Comparison Schools

We used propensity score matching to select three comparison schools to each participating ELAP school. To implement propensity score matching, we ran the following logistic regression model:

$$logit\{P(Treatment_s)\} = \beta_0 + \boldsymbol{W_s}$$

[9] VOCAL survey data are not available for K–2 schools. MA DESE also does not report VOCAL survey data for schools unless at least 10 students responded to the survey and unless the school's index score reliability was 0.7 or higher. Schools included in the sample had to have VOCAL surveys from the analytic school year (2023–24) and from the prior school year (2022–23).
[10] The sample size in Exhibit B2 may differ from the analytic sample presented later in the appendix, which includes schools for which we were able to identify matches. The matching process is described in the next section.
[11] Eligible comparison schools had to meet the following three criteria: (a) the school was categorized as a regular school by the National Center for Education Statistics (i.e., not a special education, vocational, or other/alternative school), (b) the school served students in Grades 3–8 and/or 10, and (c) school data had been reported for total enrollment, student demographics, and VOCAL survey scores in 2022–23.

where $Treatment_s$ is an indicator for whether school *s* was a treatment school (coded as 1 for ELAP schools and 0 for comparison schools); *P*($Treatment_s$) is the propensity of being a treatment school; and $\boldsymbol{W_s}$ is a vector of school-level characteristics (i.e., school locale, total student enrollment, prior-year VOCAL survey scores in Engagement, *Safety*, *Environment*, and *Overall*, and percentages of students by race/ethnicity, by gender, English learners, students with disabilities, and students from low-income households). When establishing the non-ELAP comparison group, we forced matches within school level so that participating ELAP elementary schools were matched only to other nonparticipating elementary schools; the same method was used to match middle schools with middle schools and high schools with high schools.

After matching, we assessed baseline equivalence on the matched sample to examine whether comparison schools were indeed similar to the corresponding sample of ELAP schools prior to conducting the analyses.

Exhibits B3 presents the baseline equivalence results for the matched school climate sample. The values represent the SMD between ELAP schools and matched comparison schools on their school characteristics and prior-year VOCAL survey outcomes. All SMDs for prior-year VOCAL survey outcomes were below 0.25, the required cutoff for achieving baseline equivalence.[12] *This means that any difference in outcomes is due to the presence of ELAP, not preexisting baseline differences.*

Exhibit B3. Baseline Equivalence in School Characteristics and Prior VOCAL Survey Outcomes for the School Climate Outcomes Sample

Characteristic	School climate (school *N* = 106)
Rural school	0.16
Average student enrollment	−0.09
Percentage of students who are White and non-Hispanic	0.01
Percentage of female students	−0.23
Percentage of English learners	−0.19
Percentage of students with disabilities	0.27
Percentage of students from low-income households	−0.02
Average prior-year *Overall* VOCAL score	0.01

[12] Per What Works Clearinghouse (WWC) standards, baseline equivalence is achieved with an SMD of 0.25 or below; here, between ELAP schools and comparison schools on the variables included in the matching analysis, particularly those variables related to schools' prior-year VOCAL survey outcomes. In addition, baseline characteristics for which the SMD is between 0.05 and 0.25 must be included in the analytic model as covariates to statistically adjust for any remaining baseline differences (WWC, 2022). We include all characteristics and prior-year VOCAL survey scores in our analytic models.

Characteristic	School climate (school *N* = 106)
Average prior-year *Engagement* VOCAL score	0.07
Average prior-year *Safety* VOCAL score	−0.07
Average prior-year *Environment* VOCAL score	0.04

Note. The values represent the SMD between ELAP schools and matched comparison schools.

Analyses of School Climate Outcomes

To estimate differences in VOCAL survey scores between ELAP schools and matched comparison schools, we estimated the following model separately for each outcome:

$$Y_s = \beta_0 + \beta_1 Treatment_s + \boldsymbol{W_s} + \boldsymbol{SchoolLevel_s} + v_s$$

where Y_s is the outcome of interest (i.e., VOCAL survey score)[13] for school s; $Treatment_s$ is an indicator for whether school s was a treatment school (coded as 1 for ELAP schools and 0 for comparison schools); and $\boldsymbol{W_s}$ is a vector of school-level characteristics (i.e., school locale, total student enrollment, prior-year VOCAL survey score, and percentages of students by race/ethnicity, by gender, English learners, students with disabilities, and students from low-income households). School level ($\boldsymbol{SchoolLevel_s}$) fixed effects are included to ensure that schools only within the same school level are compared with one another. Random error at the school level is captured by v_s. The main parameter of interest is β_1, which represents the estimated mean difference in the outcome measure (i.e., VOCAL survey scores) between ELAP schools and matched comparison schools.

We also conducted exploratory analyses to examine whether program impact on VOCAL survey outcomes differed by school level and school high-need status by including an interaction between the treatment and indicator of group membership and running the analytic model separately for each interaction. For example, for school level, we included the interaction term $Treatment_s * \boldsymbol{SchoolLevel_s}$ in the model. Because matching was not conducted for each individual group, these group analyses should be interpreted with caution because group membership may be associated with, or a proxy for, other factors producing differences.

Analytic Sample

The analytic sample for the study on school climate is 28 ELAP schools and 78 comparison schools. The sample is the same for all four outcomes, including VOCAL survey scores in *Engagement*, *Safety*, *Environment*, and *Overall*. The sample allows us to detect a minimum

[13] VOCAL survey scores range from 1 to 99 and are anchored on to the 2018 scale, where each scale in 2018 had a mean of 50 and a standard deviation of 20. Using benchmarks of 30, 50, and 70, the scores can be categorized into least favorable (<=30), somewhat favorable (31–50), favorable (51–70), and most favorable (>70; MA DESE, 2024).

detectable effect size of 0.22.[14] While Jacob and colleagues (2015) found an effect size on school climate of 0.45 as measured by principal surveys and 0.02 as measured by teacher surveys on the McREL Balanced Leadership program, these benchmarks are not directly comparable to the VOCAL surveys, which are completed by students. The existing literature lacks a clear benchmark for the effects of school leadership programs on school climate as measured from a student perspective.

Exhibit B4 presents the demographic characteristics for the VOCAL survey outcomes sample. The VOCAL survey sample included 106 schools, including 28 ELAP schools and 78 comparison schools. Schools in both study groups were primarily elementary schools (57% and 58%) and few were in rural settings (11% and 6%). Students in these schools were mostly White and non-Hispanic students (62% and 61%) and about one-third of the students were from low-income households (35% and 35%). Few students in these schools were English learners (8% and 11%) and about one-sixth had disabilities (18% and 17%).

Exhibit B4. School Characteristics in the VOCAL Survey Sample, by Study Group

Characteristics	ELAP schools (*N* = 28)	Comparison schools (*N* = 78)
Rural school	11%	6%
Average student enrollment	564 students	605 students
Percentage of students who are White and non-Hispanic	62%	61%
Percentage of female students	48%	49%
Percentage of English learners	8%	11%
Percentage of students with disabilities	18%	17%
Percentage of students from low-income households	35%	35%
Average prior-year *Overall* VOCAL score	52.96	52.87
Average prior-year *Engagement* VOCAL score	52.68	52.19
Average prior-year *Safety* VOCAL score	51.32	52.04
Average prior-year *Environment* VOCAL score	53.86	53.53
School level		
Elementary	57%	58%
Middle	25%	25%
High	18%	17%

[14] For school climate, the individual-level power analysis includes the following based on the analytic sample and model: α = .05; a two-tailed test; power = .80; percentage of schools assigned to treatment (ELAP) = 26%; R^2 = .84.

Results

Exhibits B5–B8 present the overall and differential impact estimates of ELAP on VOCAL survey outcomes. The estimates may be interpreted as follows.

A value of 1.39 for the *Overall* score on the VOCAL survey, for example, indicates that the average response in ELAP schools was 1.39 points higher on a 99-point scale than the average response in non-ELAP schools. This estimate is not statistically significant and suggests that the average difference in *Overall* VOCAL scores is not conclusively different between the two groups.

Exhibit B5. Impact Estimates of ELAP on School Climate—*Overall* VOCAL Score

Group	*N*	Estimate	*SE*	*p* value
Overall	106	1.39	1.09	0.20
School level				
Elementary	61	1.82	1.44	0.21
Middle	27	0.69	2.18	0.75
High	18	0.95	2.66	0.72
School high-need status [a]				
High-need	68	1.71	1.33	0.20
Not high-need	38	0.94	1.76	0.60

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.
[a] Schools with 50% or more high-need students were classified as "high need."

Exhibit B6. Impact Estimates of ELAP on School Climate—*Engagement* VOCAL Score

Group	*N*	Estimate	*SE*	*p* value
Overall	106	1.77	0.96	0.07
School level				
Elementary	61	1.94	1.27	0.13
Middle	27	2.06	1.92	0.29
High	18	0.77	2.33	0.74
School high-need status [a]				
High-need	68	1.52	1.18	0.20
Not high-need	38	2.21	1.56	0.16

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.
[a] Schools with 50% or more high-need students were classified as "high need."

Exhibit B7. Impact Estimates of ELAP on School Climate—*Safety* VOCAL Score

Group	*N*	Estimate	*SE*	*p* value
Overall	106	1.44	1.15	0.21
School level				
Elementary	61	2.27	1.51	0.14
Middle	27	−0.79	2.29	0.73
High	18	1.91	2.78	0.49
School high-need status [a]				
High-need	68	2.48	1.39	0.08
Not high-need	38	−0.10	1.86	0.96

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.
[a] Schools with 50% or more high-need students were classified as "high need."

Exhibit B8. Impact Estimates of ELAP on School Climate—*Environment* VOCAL Score

Group	*N*	Estimate	*SE*	*p* value
Overall	106	0.88	0.96	0.36
School level				
Elementary	61	1.00	1.27	0.43
Middle	27	0.52	1.93	0.79
High	18	0.98	2.35	0.68
School high-need status [a]				
High-need	68	0.88	1.17	0.45
Not high-need	38	0.93	1.54	0.55

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.
[a] Schools with 50% or more high-need students were classified as "high need."

Appendix C. School Leader Retention Study

This appendix describes the study on the impact of the Center for Leadership and Educational Equity (CLEE) Equity Leader Accelerator Program (ELAP) on school leader retention. We first describe our methodological approach, then the analytic sample, and, finally, the results.

Methodological Approach

In this section, we first describe the eligible sample and outcomes and then the analytic approach.

Eligible Sample and Outcomes

AIR measured school leader retention—both in the district and in the state—on principals and assistant principals using 2023–24 and 2024–25 data from MA DESE. We used administrative staff assignment data to determine school leaders' retention status. *Retention was defined as returning to work in a school leader position in the following school year. Based on this definition of retention, a school leader does not have to work in a district or state for a full year to be considered retained; however, they must return to a school leader position (either the same one or another role; for example, an assistant principal can come back the following school year as a principal).*

To prepare the data for analysis, AIR used the ELAP School District Crosswalk to identify ELAP-participating schools. We removed 8 schools, including those that did not report school-level data and schools that serve only K–2 students.[15] Exhibit C1 presents our identification process in the administrative data, showing how the sample size changed as we implemented these criteria.[16]

Exhibit C1. Number of Eligible ELAP Schools Included in the School Leader Retention Sample

Criteria	Retention
Number of ELAP schools in the ELAP School District Crosswalk with Early Career Principals	38
Number of ELAP schools identified in the data [a]	38
Number of ELAP schools that reported school-level data	34
Number of Grade 3 and up schools with MCAS data	30
Number of ELAP schools with a principal and/or assistant principal who was a school leader during all of the first year of implementation (2023–24)	26

[a] Not identified in the data means that an ELAP school was not found in the school-level data.

[15] MCAS student achievement data are not available for K–2 schools. Schools included in the sample had to have MCAS achievement data from the prior school year (2022–23).

[16] The sample size in Exhibit C1 may differ from the analytic sample presented later in the appendix, which includes schools for which we were able to identify matches. The matching process is described in the next section.

Analytic Approach

To examine the impact of ELAP on school leader retention, we followed a two-stage approach similar to what was described in Appendices A and B for the student achievement and school climate studies. First, we used propensity score matching to identify comparison schools[17] in the administrative data similar to ELAP schools in terms of: (a) school level (elementary, middle, or high), (b) percentages of students from low-income households, (c) MCAS proficiency rates in ELA and math in the prior school year, and (d) principal and assistant principal composition. Second, focusing solely on the schools (and the school leaders in those schools) identified in the matching process, we estimated differences in retention rates between the school leaders in ELAP schools and the school leaders in comparison schools.

Identifying Comparison Schools

We used propensity score matching to select three comparison schools to each participating ELAP school. To implement propensity score matching, we ran the following logistic regression model:

$$logit\{P(Treatment_s)\} = \beta_0 + \boldsymbol{W_s}$$

where $Treatment_s$ is an indicator for whether school *s* was a treatment school (coded as 1 for ELAP schools and 0 for comparison schools); *P*($Treatment_s$) is the propensity of being a treatment school; and $\boldsymbol{W_s}$ is a vector of school-level characteristics (i.e., percentage of students from low-income households, prior-year ELA and math proficiency rates, presence of an assistant principal, and percentages of principals and assistant principals by race/ethnicity, by gender, by novice status in their career, and by novice status in their school). When establishing the non-ELAP comparison group, we forced matches within school level so that participating ELAP elementary schools were matched only to other nonparticipating elementary schools; the same method was used to match middle schools with middle schools and high schools with high schools.

After matching, we assessed baseline equivalence on the matched sample to examine whether school leaders in comparison schools were indeed similar to the corresponding sample of school leaders in ELAP schools prior to conducting the analyses.

[17] Eligible comparison schools had to meet the following four criteria: (a) the school was categorized as a regular school by the National Center for Education Statistics (i.e., not a special education, vocational, or other/alternative school), (b) the school served students in Grades 3–8 and/or 10, (c) school data had been reported for school level, locale, the proportion of students from low-income households, and MCAS scores in 2022–23, and (d) principal and/or assistant principal data had been reported in 2023–24 MA DESE educator data.

Exhibit C2 presents the baseline equivalence results at the school and school leader levels for the matched retention sample.[18] The values represent the SMD between school leaders in ELAP schools and school leaders in matched comparison schools on their demographics and school characteristics, including prior-year achievement from the state assessments. All SMDs for prior-year achievement, years of school leader experience (percentage novice), and degree of disadvantage (percentage of low-income students) were below 0.25, the required cutoff for achieving baseline equivalence.[19] *This means that any difference in outcomes is due to the presence of ELAP, not preexisting baseline differences.*

Exhibit C2. Baseline Equivalence in School and School Leader Characteristics and Prior Achievement for the School Leader Retention Outcomes Sample

Characteristic	Retention (school *N* = 110; school leader *N* = 217)
School characteristics	
Presence of an assistant principal	0.12
Percentage of female principals	−0.16
Percentage of principals who were novice principals in their career (i.e., with 0 to 3 years of experience as a principal in Massachusetts in 2022–23)	−0.22
Percentage of principals who were novice principals in their school (i.e., with 0 to 3 years of experience as a principal in their school in2022–23)	−0.10
Percentage of principals who are White and non-Hispanic	−0.30
Percentage of female assistant principals	0.35
Percentage of assistant principals who were novice principals in their career (i.e., with 0 to 3 years of experience as a principal in Massachusetts in 2022–23)	0.07
Percentage of assistant principals who were novice principals in their school (i.e., with 0 to 3 years of experience as a principal in their school in2022–23)	0.01
Percentage of assistant principals who are White and non-Hispanic	−0.09
Percentage of students from low-income households	0.12
Average prior-year ELA proficiency rate	−0.09
Average prior-year math proficiency rate	−0.03

[18] Although matching took place at the school level, we also assessed whether the matches were adequate at the school leader level (i.e., on school leader-level variables, including demographics and years of experience) for the retention analyses.
[19] Per What Works Clearinghouse (WWC) standards, baseline equivalence is achieved with an SMD of 0.25 or below; here, between school leaders in ELAP schools and school leaders in comparison schools on the variables included in the matching analysis, particularly those variables related to prior-year achievement, years of school leader experience, and degree of disadvantage. In addition, baseline characteristics for which the SMD is between 0.05 and 0.25 must be included in the analytic model as covariates to statistically adjust for any remaining baseline differences (WWC, 2022). We include all characteristics and prior-year achievement rates in our analytic models.

Characteristic	Retention (school *N* = 110; school leader *N* = 217)
School leader characteristics	
Female	0.12
Novice in their career (i.e., with 0 to 3 years of experience in their current position in Massachusetts in 2022–23)	0.04
Novice in their school (i.e., with 0 to 3 years of experience in their current position in their school in 2022–23)	0.08
School leaders who are White and non-Hispanic	−0.13

Note. The values represent the SMD between school leaders in ELAP schools and school leaders in matched comparison schools.

Analyses of School Leader Retention Outcomes

To estimate differences in school leader retention rates between the school leaders in ELAP schools and the school leaders in matched comparison schools, we estimated the following logistic regression model separately for each outcome:[20]

$$Y_j = \beta_0 + \beta_1 Treatment_j + \boldsymbol{Z_j} + \boldsymbol{W_s} + \boldsymbol{SchoolLevel_s} + u_j$$

where Y_j is the outcome of interest (i.e., school leader retention at the district level, coded as 1 for retained in a school leader position within the district and 0 otherwise; or school leader retention at the state level, coded as 1 for retained in a school leader position within the state and 0 otherwise) for school leader *j*; $Treatment_j$ is an indicator for whether a school leader was in a treatment school (coded as 1 for working in an ELAP school and 0 for working in a comparison school); $\boldsymbol{Z_j}$ is a vector of school leader characteristics (i.e., race, gender, novice status in career, and novice status in school);[21] and $\boldsymbol{W_s}$ is a vector of school-level characteristics (i.e., school locale, total student enrollment, school prior-year student achievement meeting-expectations status, and percentages of students by race/ethnicity, by gender, and students from low-income households). School level ($\boldsymbol{SchoolLevel_s}$) fixed effects are included to ensure that schools only within the same school level are compared with one another. Random error at the school leader level is captured by u_j. The main parameter of interest is β_1, which represents the estimated mean difference in the outcome measure (i.e., retention rates) between school leaders in ELAP schools and school leaders in matched comparison schools.

We also conducted exploratory analyses to examine whether program impact on school leader retention outcomes differed by school level, school high-need status, school prior-year student

[20] Although school leaders were nested within schools, we estimated single-level models because there were too few school leaders per school (1.97 school leaders per school, on average) to estimate multilevel models.
[21] We were unable to control for position in the analytic model because over 95% of school leaders in the data were principals.

achievement meeting-expectations status, and novice in career status by including an interaction between the treatment and indicator of group membership and running the analytic model separately for each interaction.[22,23] For example, for school level, we included the interaction term $Treatment_s * \boldsymbol{SchoolLevel_s}$ in the model. Because matching was not conducted for each individual group, these group analyses should be interpreted with caution because group membership may be associated with, or a proxy for, other factors producing differences.

To examine program impact on retention outcomes for all staff who were school leaders during some or all of the first year of ELAP implementation (2023–24), we conducted exploratory analyses using a broader sample that includes leaders who transitioned into district leadership roles during that year. By expanding our sample, we can examine whether ELAP impacts role-to-role retention and retention as school leaders transition into district leadership roles.

Transforming Effects

To facilitate interpretation, all treatment effect estimates for retention outcomes were transformed into the odds ratio metric using the following formula:

$$Odds\ ratio = e^{LOR}$$

where e^{LOR} is the exponential of the treatment effect estimate from the logistic regression (or log odds). The odds ratio tells us how much more likely school leaders in ELAP schools are to be retained within the district or state than school leaders in comparison schools.

Analytic Sample

The analytic sample for the study on school leader retention is 58 school leaders in 26 ELAP schools and 159 school leaders in 84 comparison schools. The sample is the same for both outcomes, including retention at the district level and retention at the state level. The sample allows us to detect a minimum detectable effect size of 0.35.[24] This demonstrates our study is well-powered and sufficient to detect effects that are at or below the magnitude of those typically found for school leader retention in the literature. For example, a study on the McREL Balanced Leadership program, which was a randomized controlled trial that met WWC

[22] We also planned to examine whether program impact on school leader retention outcomes differed by race and position; however, we were unable to conduct these analyses because more than 90% of school leaders in the data were White and non-Hispanic and over 95% of school leaders were principals.

[23] Due to nearly all school leaders in the treatment group being retained, the models for the school level and school prior-year student achievement meeting-expectations status analyses had quasi-complete separation. To address this issue, we applied a Firth correction.

[24] For school leader retention, the individual-level power analysis includes the following based on the analytic sample and model: α = .05; a two-tailed test; power = .80; percentage of school leaders assigned to treatment (ELAP) = 27%; R^2 = .14.

standards, found an effect size of 0.45 on school leader retention at the school level (Jacob et al., 2015; WWC, 2020).

Exhibit C3 presents the demographic characteristics for the school leader retention sample. The retention sample included 217 school leaders, including 58 school leaders in 26 ELAP schools and 159 school leaders in 84 comparison schools. A little over half of the school leaders in both study groups were female (59% and 53%), most were White and non-Hispanic (90% and 93%), a little less than half were novice in their current position in their career (43% and 41%), and more than half were novice in their current position in their school (64% and 60%). In addition, about half of the school leaders were in elementary schools (50% and 52%), and in schools where about one-third of the students were from low-income households (37% and 34%) and about half of the students were proficient in ELA (50% and 52%) or math (48% and 48%).

Exhibit C3. School Leader and School Characteristics in the School Leader Retention Sample, by Study Group

Characteristics	School leaders in ELAP schools (SL *N* = 58; school *N* = 26)	School leaders in comparison schools (SL *N* = 159; school *N* = 84)
School leader characteristics		
Female	59%	53%
Novice in their career (i.e., with 0 to 3 years of experience in their current position in Massachusetts in 2022–23)	43%	41%
Novice in their school (i.e., with 0 to 3 years of experience in their current position in their school in 2022–23)	64%	60%
School leaders who are White and non-Hispanic	90%	93%
School characteristics		
Presence of an assistant principal	97%	94%
Percentage of female principals	38%	46%
Percentage of principals who were novice principals in their career (i.e., with 0 to 3 years of experience as a principal in Massachusetts in 2022–23)	24%	35%
Percentage of principals who were novice principals in their school (i.e., with 0 to 3 years of experience as a principal in their school in 2022–23)	45%	50%
Percentage of principals who are White and non-Hispanic	83%	92%
Percentage of female assistant principals	69%	54%
Percentage of assistant principals who were novice principals in their career (i.e., with 0 to 3 years of experience as a principal in Massachusetts in 2022–23)	47%	43%

Characteristics	School leaders in ELAP schools (SL *N* = 58; school *N* = 26)	School leaders in comparison schools (SL *N* = 159; school *N* = 84)
Percentage of assistant principals who were novice principals in their school (i.e., with 0 to 3 years of experience as a principal in their school in 2022–23)	66%	66%
Percentage of assistant principals who are White and non-Hispanic	84%	87%
Percentage of students from low-income households	37%	34%
Average prior-year ELA proficiency rate	50%	52%
Average prior-year math proficiency rate	48%	48%
School level		
Elementary	50%	52%
Middle	31%	28%
High	19%	20%

Results

Exhibit C4 presents the overall and differential impact estimates of ELAP on school leader retention in the district. We did not conduct state-level retention analyses because all school leaders in both ELAP and comparison schools were retained in the state (i.e., state-level retention was 100% for both groups). The odds ratios (last column) may be interpreted as follows.

A value of 3.74 for retention, for example, indicates that school leaders in ELAP schools were 3.74 more times likely to be retained within the district as a school leader than school leaders in comparison schools.

Exhibit C4. Impact Estimates of ELAP on School Leader Retention in the District

Group	*N*	Log odds	*SE*	*p* value	Odds ratio
Overall	217	1.32	0.53	0.01	3.74
School level					
Elementary	100	4.53	1.38	< 0.01	92.54
Middle	64	1.75	0.63	0.02	5.73
High	53	−2.01	1.09	0.07	0.13
School high-need status [a]					
High-need	81	2.83	1.16	0.01	16.90
Not high-need	136	0.35	0.77	0.65	1.42

Group	*N*	Log odds	*SE*	*p* value	Odds ratio
School prior-year student achievement meeting-expectations status [b]					
Meeting-expectations	104	6.38	1.72	< 0.01	589.74
Not meeting expectations	113	0.02	0.62	0.98	1.02
Novice in career status [c]					
Novice	90	1.86	0.64	< 0.01	6.39
Not novice	127	0.42	0.83	0.61	1.53

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.
[a] Schools with 50% or more high-need students were classified as "high need." [b] Schools with 50% or more students who met or exceeded expectations on the prior-year ELA and math state assessments were classified as "meeting expectations." [c] Novice school leaders were defined as those with 0 to 3 years of experience in their current position in Massachusetts at baseline (i.e., in 2022–23).

Exhibits C5 and C6 present the overall and differential impact estimates of ELAP on school or district leader retention in the district and in the state using our broader sample that includes all staff who were school leaders during some or all of the first year of ELAP implementation (2023–24), which allows for the inclusion of school leaders who transitioned into district leadership roles in that year. It also allowed us to conduct state-level retention analyses, as not all school or district leaders were retained in the broader sample. The estimates may be interpreted as follows.

A value of 0.92 for retention, for example, indicates that school or district leaders in ELAP schools were 0.08 less times likely to be retained within the district as a school or district leader than school or district leaders in comparison schools. However, this estimate is not statistically significant and suggests that the likelihood of retention in the district is not conclusively different between the two groups.

The analyses on the broader sample show a somewhat different pattern of results than the sample that only includes school leaders who were school leaders during all of 2023–24. For district-level retention, the odds ratio is no longer statistically significant and changes from 3.74 in the main sample to 0.92 in the broader sample. Additionally, the odds ratios for school or district leaders in elementary schools (main sample = 92.54; broader sample = 1.20), for school or district leaders in schools in which 50% or more students were meeting achievement expectations in the prior year (main sample = 589.74; broader sample = 0.71), and for school or district leaders who were novice in their career (main sample = 6.39; broader sample = 2.08) are no longer statistically significant. On the other hand, the odds ratios for school or district leaders in high schools (main sample = 0.13; broader sample = 0.05) and school or district leaders in non-high-need schools (main sample = 1.42; broader sample = 0.35) were not statistically significant in the main sample but are significant and negative in the broader sample.

For state-level retention in the broader sample, we found that school or district leaders in ELAP schools were about half as likely to be retained within the state as a school or district leader than school or district leaders in comparison schools (odds ratio = 0.50). The pattern of results differed for different groups of schools and school or district leaders. When compared with school or district leaders in comparison schools, retention was statistically significantly higher for school or district leaders in ELAP schools in middle schools (odds ratio = 8.36) but significantly lower for school or district leaders in ELAP schools in elementary (odds ratio = 0.34) and high (odds ratio = 0.07) schools, in non-high-need schools (odds ratio = 0.14), in schools in which 50% or more students were meeting achievement expectations in the prior year (odds ratio = 0.18), and for school or district leaders who were not novice in their career (odds ratio = 0.28).

These results suggest that ELAP does not mitigate retention losses related to transitions from school leadership to district leadership roles that occurred during the first year of program implementation.

Exhibit C5. Impact Estimates of ELAP on School or District Leader Retention in the District—Broader Sample

Group	*N*	Log odds	*SE*	*p* value	Odds ratio
Overall	262	−0.08	0.25	0.75	0.92
School level					
Elementary	118	0.18	0.39	0.64	1.20
Middle	81	2.13	0.44	< 0.01	8.44
High	63	−2.96	0.57	< 0.01	0.05
School high-need status [a]					
High-need	99	1.32	0.44	< 0.01	3.74
Not high-need	163	−1.06	0.36	< 0.01	0.35
School prior-year student achievement meeting-expectations status [b]					
Meeting-expectations	122	−0.34	0.40	0.39	0.71
Not Meeting expectations	140	0.11	0.33	0.73	1.12
Novice in career status [c]					
Novice	109	0.73	0.39	0.06	2.08
Not novice	153	−0.59	0.33	0.08	0.56

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.
[a] Schools with 50% or more high-need students were classified as "high need." [b] Schools with 50% or more students who met or exceeded expectations on the prior-year ELA and math state assessments were classified as "meeting expectations." [c] Novice school leaders were defined as those with 0 to 3 years of experience in their current position in Massachusetts at baseline (i.e., in 2022–23).

Exhibit C6. Impact Estimates of ELAP on School or District Leader Retention in the State—Broader Sample

Group	*N*	Log odds	*SE*	*p* value	Odds ratio
Overall	262	−0.69	0.30	0.02	0.50
School level					
Elementary	118	−1.07	0.47	0.02	0.34
Middle	81	2.12	0.55	< 0.01	8.36
High	63	−2.62	0.60	< 0.01	0.07
School high-need status [a]					
High-need	99	0.78	0.53	0.14	2.19
Not high-need	163	−1.95	0.45	< 0.01	0.14
School prior-year student achievement meeting-expectations status [b]					
Meeting-expectations	122	−1.73	0.53	< 0.01	0.18
Not Meeting expectations	140	< 0.01	0.39	1.00	1.00
Novice in career status [c]					
Novice	109	0.21	0.47	0.65	1.24
Not novice	153	−1.27	0.40	< 0.01	0.28

Note. SE = standard error. Estimates with *p* values less than .05 are considered statistically significant.
[a] Schools with 50% or more high-need students were classified as "high need." [b] Schools with 50% or more students who met or exceeded expectations on the prior-year ELA and math state assessments were classified as "meeting expectations." [c] Novice school leaders were defined as those with 0 to 3 years of experience in their current position in Massachusetts at baseline (i.e., in 2022–23).

References

Hedges, L. V., & Olkin, I. (2016). Overlap between treatment and control group distributions of an experiment as an effect size measure. *Psychological Methods, 21*(1), 61–68. https://doi.org/10.1037/met0000042

Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2015). Exploring the causal impact of the McREL Balanced Leadership program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis, 37*(3), 314–332. https://eric.ed.gov/?id=EJ1072764

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Massachusetts Department of Elementary and Secondary Education. (2022). *2022 Massachusetts state report card.* https://profiles.doe.mass.edu/staterc/?fyCode=2022

Massachusetts Department of Elementary and Secondary Education. (2024). *Views of climate and learning (VOCAL) survey project.* https://www.doe.mass.edu/research/vocal/

Massachusetts Department of Elementary and Secondary Education. (n.d.). *Profiles help—About the data: Public schools and districts.* https://profiles.doe.mass.edu/help/data.aspx?section=students

Massachusetts Department of Elementary and Secondary Education, Office of Planning and Research. (2024). *Researcher's guide to Massachusetts state education data.* https://www.doe.mass.edu/research/researchers.html

What Works Clearinghouse. (2020, March). *Balanced Leadership®*. U.S. Department of Education, Institute of Education Sciences. https://whatworks.ed.go

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, Version 5.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf

About the American Institutes for Research®

Established in 1946, the American Institutes for Research ® (AIR ®) is a nonpartisan, not-for-profit institution that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of education, health, and the workforce. AIR's work is driven by its mission to generate and use rigorous evidence that contributes to a better, more equitable world. With headquarters in Arlington, Virginia, AIR has offices across the U.S. and abroad. For more information, visit AIR.ORG

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

28745_09/25