

Study Brief

Impact Evaluation of the Equity Leader Accelerator Program: Year 1 Findings

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

September 2025

The American Institutes for Research[®] (AIR[®]) evaluated the impact of the Center for Leadership and Educational Equity (CLEE) Equity Leader Accelerator Program (ELAP) during the 2023–24 school year. ELAP aims to strengthen the equity-centered leadership practices of school leaders in each of four career stages—Emerging Leaders, Aspiring Principals, Early Career Principals, and Experienced Leaders. This brief summarizes the impact of ELAP on student academic achievement, school climate, and school leader retention after 1 year of implementation in Massachusetts. Technical appendices provide additional details.

How was the program implemented?

ELAP provides 2-years of differentiated, equity-centered leadership development for educators at four career stages: (1) Emerging Leaders, which includes early-career teacher leaders beginning to explore formal leadership pathways; (2) Aspiring Principals, which includes teacher leaders preparing to earn administrator certification; (3) Early Career Principals, which includes principals or assistant principals in their first 5 years of school leadership; and (4) Experienced Leaders, which includes school or district administrators with more than 5 years of leadership experience. Leaders are trained to implement CLEE’s six Core Leadership Practices (Setting Direction, Monitoring Progress, Building Capacity to Teach, Building Capacity to Collaborate, Building Capacity to Lead, and Reorganizing Systems) to facilitate school improvement through practitioner-based improvement cycles.

The program included two primary components in which participants were expected to engage: job-embedded authentic practice (i.e., at least 20 mentoring meetings) and eight 60-minute one-on-one coaching sessions with two observations. In practice, most participants engaged in fewer than the expected number of mentoring meetings (25% of participants meeting 14 or fewer times, 6% meeting 15–17 times, and 69% meeting 18 or more times), while the majority met or exceeded expectations for coaching sessions (87% completing 7–8 sessions) and observations (63% completing both observations). The program also included two optional components: communities of practice (i.e., networking and peer learning sessions) and asynchronous learning modules. These components were delivered primarily in a virtual format, with some in-person coaching and mentoring sessions depending on participant and coach availability.

ELAP emphasized sustained, job-embedded learning and cross-role mentorship within a networked learning community. In Year 1 (2023–24), Early Career Principals led school improvement efforts aligned to equity priorities while receiving mentoring from Experienced Leaders; both groups also received CLEE coaching, accessed asynchronous Canvas modules, and participated in virtual peer learning and networking sessions.

What implementation challenges emerged?

Implementation challenges emerged in Year 1 related to recruitment, clarity of expectations, and district coordination, as indicated by interviews conducted by AIR with ELAP participants and staff (see Acheampong et al., 2024). In some cases, participants entered the program without a clear understanding of its purpose or the expectations for their engagement, which contributed to early disengagement.

Communication gaps between district leaders and school-level participants, along with leadership turnover, contributed to a disruption in continuity and support, which led to inconsistent implementation. Some participants also perceived ELAP requirements as an additional burden rather than an integrated form of support, especially during the demanding start of the school year. They reported that program assignments often duplicated existing tasks and were difficult to integrate with other responsibilities, such as budgeting and routine school management.

In response to this feedback, CLEE made several adjustments for Year 2, including clearer onboarding materials, more proactive district engagement, simplified tools, and improved communication strategies.

What were the program impacts?

AIR's impact evaluation consisted of three studies conducted in Massachusetts. The first study examined the impact of ELAP on student achievement in English language arts (ELA) and math in Grades 4–8¹ in schools with participating early career principals. The second study examined the impact of ELAP on school climate in Grades 4, 5, 8, and 10, as measured by the

Views of Climate and Learning (VOCAL) survey, in schools of participating early career principals.²

The third study examined the impact of ELAP on retention of school leaders, specifically principals and assistant principals. AIR used a quasi-experimental design based on samples constructed through propensity score matching in all three studies.³

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.

¹ Massachusetts Comprehensive Assessment System (MCAS) state assessments are administered in Grades 3–8 and 10; thus, student achievement analyses were restricted to Grades 4–8 because prior-year scores are not available for students in Grades 3 and 10 for matching purposes.

² The VOCAL survey measures students' perceptions of overall school climate and the dimensions of *Engagement*, *Safety*, and *Environment*. Students in Grades 4, 5, 8, and 10 can participate in the survey as part of the state MCAS administration each year.

³ All three studies were designed to generate rigorous evidence that will meet What Works Clearinghouse (WWC) evidence standards with reservations. We employed propensity score matching to estimate the predicted probability of group membership (treatment versus comparison school) based on observed school characteristics and baseline achievement scores aggregated to the school level for student achievement outcomes, baseline VOCAL survey scores aggregated to the school level for school climate outcomes, or school leader characteristics for the school leader retention outcome. We then used that predicted probability to create a comparison group similar to the treatment group on these observed characteristics and baseline outcomes (Hansen, 2004). Last, we used regression models to estimate program impacts on student and school leader outcomes, controlling for school characteristics, baseline outcomes, and student or school leader demographics. We present findings only for those outcomes and samples on which we achieved baseline equivalence. Per WWC standards, we defined the term “baseline equivalence” as obtaining a standardized mean difference of 0.25 or below between ELAP schools and matched comparison schools on baseline outcomes (achievement scores at the student level for student achievement outcomes; VOCAL survey scores at the school level for school climate outcomes; and school leader experience, degree of disadvantage, and student achievement scores at the school level for school leader retention outcomes). See the technical appendices for additional details on the methodological approach.

The impact evaluation included the following samples:

- In the study on the impact of ELAP on student achievement, the ELA student sample included 7,278 students in 25 ELAP schools and 18,286 students in 74 comparison schools. The math student sample included 7,318 students in 25 ELAP schools and 18,345 students in 74 comparison schools.
- In the study on the impact of ELAP on school climate, the VOCAL survey sample included 28 ELAP schools and 78 comparison schools.
- In the study on the impact of ELAP on school leader retention, the sample included 58 school leaders in 26 ELAP schools and 159 school leaders in 84 comparison schools.

Students in ELAP and non-ELAP schools had similar spring 2024 ELA and math scores, on average; results for focal students in ELAP schools were mixed.

Achievement scores for students in ELAP schools in 2023–24 did not differ significantly from those of similar students in comparison schools. The overall estimated mean difference in ELA achievement scores was -0.02 *standard deviations (SD)*. Based on the What Works Clearinghouse (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.⁴ However, the difference was not statistically significant. The overall estimated mean difference in math achievement scores was 0.02 *SD*, which corresponds to a WWC improvement index of $+0.88$ for students in ELAP schools. The difference in math scores also was not statistically significant. Together, these findings indicate that ELAP did not have a significant impact on ELA or math achievement for students overall.

The pattern of results for different groups of schools and students⁵ was generally consistent with the overall results, except for middle school grades for ELA achievement and elementary school grades for math achievement.⁶ When compared with students in non-ELAP schools, ELA scores were significantly lower for

Summary of Student Achievement Findings

Study Sample	Student ELA Achievement	Student Math Achievement
Overall		
Elementary school grades		
Middle school grades		
High-need students		
Not high-need students		
Focal students *		

* The analysis on focal students is not based on the quasi-experimental design, as these students were not matched.

 ELA – not significant at the 0.05 level
  MATH – not significant at the 0.05 level
 ELA – statistically significant and positive effect of ELAP
  MATH – statistically significant and positive effect of ELAP
 ELA – statistically significant and negative effect of ELAP
  MATH – statistically significant and negative effect of ELAP

⁴ For achievement, we transformed the estimated effect sizes into the WWC improvement index metric for interpretation. 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 (Hedges & Olkin, 2016; WWC, 2022). See the technical appendices for additional guidance on interpretation.

⁵ We examined whether differences in student achievement outcomes between students in ELAP schools and students in comparison schools varied by school level and student high-need status. MA DESE (2024) defines “high need” students as those from low-income households, English learners (current or former), or students with disabilities.

⁶ Elementary school grades include Grades 4–5 and middle school grades include Grades 6–8.

students in ELAP schools in middle school grades ($SD = -0.08$; WWC improvement index = -3.14) and math scores were significantly higher for students in ELAP schools in elementary school grades ($SD = 0.06$; WWC improvement index = $+2.52$).

To examine program performance on “focal” students, we also examined differences in outcomes within ELAP schools between focal students and nonfocal students.⁷ We found significant differences between focal and nonfocal students on measures of both ELA and math achievement. When compared with nonfocal students in ELAP schools, focal students in ELAP schools had significantly higher ELA scores ($SD = 0.05$; WWC improvement index = $+2.04$) but significantly lower math scores ($SD = -0.05$; WWC improvement index = -2.14). However, these findings should be interpreted with caution as focal and nonfocal students were not matched and differ at baseline by design (or definition).

Students in ELAP schools, on average, reported comparable perceptions on school climate outcomes in 2023–24 compared with students in non-ELAP schools.

Overall student perceptions on school climate did not differ between ELAP schools and similar comparison schools in 2023–24. On average, students in ELAP schools responded 1.39 points higher on a 99-point scale on the *Overall* VOCAL survey than students in comparison schools. The VOCAL survey includes three dimensions: (1) *Engagement*, which includes measures of cultural competence, participation, and relationships; (2) *Safety*, which includes measures of emotional safety, physical safety, and bullying; and (3) *Environment*, which measures instructional environment, mental health, and discipline environment. Students in ELAP schools responded 1.77 points higher on *Engagement*, 1.44 points higher on *Safety*, and 0.88 points higher on *Environment*. However, the differences were not statistically significant for any of the VOCAL survey outcomes. These results indicate that, on average, both ELAP and non-ELAP schools had similar perceptions on school climate outcomes; both groups also met the “favorable” school climate benchmark.⁸

The pattern of results for different groups of schools⁹ was consistent with the overall results.

Summary of School Climate Findings

Study Sample	Overall VOCAL Score	Engagement VOCAL Score	Safety VOCAL Score	Environment VOCAL Score
Overall				
Elementary schools				
Middle schools				
High schools				
High-need schools				
Not high-need schools				

 *Overall* – not significant at the 0.05 level
  *Safety* – not significant at the 0.05 level
  *Engagement* – not significant at the 0.05 level
  *Environment* – not significant at the 0.05 level

⁷ 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.

⁸ 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).

⁹ We examined whether differences in school climate outcomes between ELAP schools and comparison schools varied by school level and school high-need status. Schools with 50% or more high-need students were classified as “high need.”

School leaders in ELAP schools, on average, were more likely to be retained in the district than school leaders in non-ELAP schools.

School leaders in ELAP schools were 3.74 times more likely to be retained¹⁰ within the district than school leaders in non-ELAP schools (odds ratio = 3.74, $p = 0.01$).^{11,12} The pattern of results for different groups of schools and school leaders¹³ was generally consistent with the overall results. When compared with school leaders in non-ELAP schools, retention was significantly higher for school leaders in ELAP schools in elementary (odds ratio = 92.54, $p < 0.01$) and middle (odds ratio = 5.73, $p = 0.02$) schools; in high-need schools (odds ratio = 16.90, $p = 0.01$); in schools in which 50% or more students were meeting achievement expectations in the prior year (odds ratio = 589.74, $p < 0.01$); and for school leaders who were novice in their career (odds ratio = 6.39, $p < 0.01$).

Implications

After 1 year of implementation, ELAP had a mix of positive, negative, and null findings. Overall, the program shows great promise in school leader retention. The program's success in retaining school leaders, particularly those who are new to their roles and in high-need schools, is a strong indicator that ELAP is an effective tool for talent development and retention within the district. There were also positive impacts of ELAP on math achievement at the elementary school level and for focal students' ELA achievement.

Null findings for school climate and student achievement may be due to a mixture of factors. One possibility is that 1 year of implementation is not sufficient to observe significant differences between ELAP and comparison schools for student outcomes. This is not uncommon for a new program, as changes to student

Summary of School Leader Retention Findings

Study Sample	School Leader Retention in the District
Overall	
Elementary schools	
Middle schools	
High schools	
High-need schools	
Not high-need schools	
Meeting achievement expectations schools	
Not meeting achievement expectations schools	
Novice in career school leaders	
Not novice in career school leaders	

 Retention – not significant at the 0.05 level

 Retention – statistically significant and positive effect of ELAP

¹⁰ We defined retention 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).

¹¹ For retention, we transformed the estimate from the logistic regression (or log odds) into the odds ratio metric for interpretation. The odds ratio tells us how much more likely school leaders in ELAP schools are to be retained within the district than school leaders in comparison schools.

¹² 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).

¹³ We examined whether differences in retention outcomes between school leaders in ELAP schools and school leaders in comparison schools varied by school level, school high-need status, school prior-year student achievement meeting-expectations status, and novice in career status. Schools with 50% or more high-need students were classified as "high need." 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." Novice school leaders were defined as those with 0 to 3 years of experience in their current position in Massachusetts.

outcomes typically take more time to manifest. A second possible explanation is that the measures of school climate used in the VOCAL survey are not aligned with the specific outcomes the ELAP program seeks to improve. Third, the lack of measurable effects on these outcomes could be due to strong existing practices in the comparison schools. For example, in elementary settings, null findings in ELA may be due to Massachusetts' heavy investment in evidence-based practices and a statewide push of Multi-Tiered Systems of Support (MTSS) in literacy.¹⁴ This initiative aims to identify and support students who need reading interventions to ensure all students receive a high level of support. This means that students in both ELAP and comparison schools may be receiving similar, high-quality interventions, which would make it difficult to detect a unique effect from the ELAP program.

Finally, the negative findings for ELA achievement in middle school grades and math achievement for focal students warrant further investigation to understand the mechanisms and specific contextual factors at play. These results may be influenced by differences in school structure, where elementary schools can dedicate more time to the program.

As these are preliminary findings from the first year of a 2-year program, they serve as a valuable tool for future program development and for understanding how the initiative might be adapted to other contexts. The full 2-year evaluation will provide a more comprehensive picture of ELAP's overall effects.

Limitations

The results presented here should be interpreted with caution due to several limitations:

- **Lack of random assignment.** The studies in the impact evaluation relied on a quasi-experimental design rather than a randomized controlled trial, in which schools are randomly assigned to participate in ELAP or not. ELAP schools and comparison schools were equivalent on baseline observable characteristics, including prior test scores of students. However, because there was no random assignment, it is possible that ELAP schools and comparison schools were different at baseline on characteristics that were not measured (e.g., motivation). Given the quasi-experimental design, we cannot rule out the possibility that differences in the outcome measures between ELAP and non-ELAP schools are due to unobserved baseline differences between these two study conditions.
- **Sample representativeness.** Due to baseline data requirements, the study samples represent a fraction of schools participating in ELAP. The study on student achievement relies on prior-year student-level achievement scores to assess baseline equivalence; thus, the sample is restricted to students in tested grades with prior-year scores (Grades 4–8). This represents about 66% of schools participating in ELAP. Similarly, the school climate study relies on prior-year school-level VOCAL survey scores, and the school leader retention study relies on prior-year school-level achievement scores, restricting both studies to schools that serve Grades 3 and up. The school climate study sample represents 74% of schools

¹⁴ The Multi-Tiered System of Support (MTSS) is a proactive and preventative framework designed to ensure equitable access to grade-level learning and beyond for all students. MTSS organizes evidence-based practices across three tiers of support—universal (Tier 1), targeted (Tier 2), and intensive (Tier 3)—to address academic, social-emotional, behavioral, and wellness needs. The Massachusetts MTSS Blueprint emphasizes inclusive, culturally responsive instruction, data-driven decision making, and strategic resource alignment to close persistent opportunity gaps. When applied to literacy, MTSS-R integrates structured literacy instruction with differentiated interventions in reading. The Massachusetts Department of Elementary and Secondary Education (DESE) has launched initiatives such as the Massachusetts Literacy Institute to build organizational capacity for MTSS-R implementation. These efforts include professional development, coaching, and tools to help educators deliver high-quality, evidence-based literacy instruction aligned with the Science of Reading.

participating in ELAP and the school leader study sample represents 68% of schools participating in ELAP. These sample restrictions suggest the findings may not be generalizable to all ELAP schools.

- **Implementation challenges.** These results represent impacts of ELAP after only 1 year of implementation of the 2-year program. During that year, ELAP experienced challenges related to communication and leadership turnover, which may have affected implementation. These implementation challenges suggest the findings may not be generalizable to a typical year of implementing ELAP.

Acknowledgments

The project was conducted as part of the 2022 Supporting Effective Educator Development (SEED) Grant independent evaluation requirement under award number S423A220015. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Department of Education.

We would like to extend our sincere appreciation to all members of the project team whose dedication and contributions were instrumental to the success of this research study. They include Ms. Abena Acheampong, Ms. Melissa Arellanes, Mr. Peter Coyne, Mr. Kendell Daughtry, Ms. Jasmine James, Dr. Angela Johnson, Dr. Eric Larson, Dr. Stephanie Levin, and Dr. Burhan Ogut. Although not named as key authors, their efforts in data collection, analysis, coordination, quality assurance and operational support played a vital role in making this study possible. Their commitment and collaboration are deeply valued and gratefully acknowledged.

References

- Acheampong, A., Arellanes, M., Brown-Sims, M., Coyne, P., James, J., Johnson, A., Larsen, E., & Olivar, S. M. (2024, September). *Equity Leader Accelerator Program Evaluation: Analysis of Year 1 of Implementation (Year 2 Biannual Brief 2)*. Center for Leadership and Educational Equity.
- Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99(467), 609–618.
<https://doi.org/10.1198/016214504000000647>
- 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>
- Massachusetts Department of Elementary and Secondary Education. (n.d.). *Profiles help—About the data: Public schools & districts*. <https://profiles.doe.mass.edu/help/data.aspx?section=students>
- 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, 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. (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



Notice of Trademark: "American Institutes for Research" and "AIR" are registered trademarks. All other brand, product, or company names are trademarks or registered trademarks of their respective owners.
Copyright © 2025 American Institutes for Research®. All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, website display, or other electronic or mechanical methods, without the prior written permission of the American Institutes for Research. For permission requests, please use the Contact Us form on www.air.org.