Significance

Summary of Innovation

The Massachusetts Department of Elementary and Secondary Education (DESE) proposes to implement and study the effectiveness of an early-grade literacy innovation that combines two interventions: 1) high-dosage, evidence-based tutoring in 1st grade, and 2) follow-up lighter-dosage, evidence-based tutoring in 2nd grade. "BOOST: Building Ongoing Outcomes in Student Tutoring" aims to answer critical questions about how gains from high-dosage literacy tutoring (HDLT) can be maximized and sustained over time. Our proposed study will generate new evidence on the conditions and contexts for sustained impact of high-dosage tutoring over time and the cost-effectiveness of adding low dosage literacy tutoring (LDLT) to sustain student gains from high-dosage tutoring beyond the initial treatment period. Massachusetts recognizes the critical nature of early literacy as the gateway to future academic success and seeks to further invest in, refine and scale high-impact tutoring, a proven intervention.

Intervention 1: Deliver High-Dosage Literacy Tutoring in 1st Grade

High-dosage tutoring is an early literacy intervention with robust evidence of effectiveness in a range of contexts (Foorman, et al, 2016; Tepper Jacob, et al, 2015; Markovitz, et al, 2014). BOOST will build on Massachusetts' successful statewide Early Literacy Tutoring Program, which launched in 2021 with 12 local educational agencies (LEAs) and 1,201 students to mitigate the impact of disrupted learning due to the COVID pandemic. The program has steadily grown over the past four years, and in SY2025-2026 will serve approximately 66 LEAs, 325 schools and 10,000 students. The existing program focuses on <u>high-dosage</u> tutoring offered in 10-12 week cycles with 2-5 sessions per week delivered by state-approved vendors of <u>evidence-</u>

based high-impact literacy tutoring models.[1] Tutoring is delivered to high need students in grades K-3, which DESE defines as students who score below benchmark on a state-approved early literacy screening assessment during the same academic year as they receive tutoring.

In 2022, DESE contracted with American Institutes for Research (AIR) to evaluate the Early Literacy Tutoring Program using a quasi-experimental design (see Appendix J for 2023 and 2024 reports). Figure 1, below, presents the scale and dosage of the DESE Early Literacy Tutoring Program included in the AIR study.

Figure 1: Scale & Dosage of DESE Early Literacy Tutoring Program

	Scale	**Dosage**
SY 2022-2023	- 4 vendors - 14 LEAs - 67 schools - 2,907 students	- 89,755 tutoring sessions - 37.97 tutoring sessions per student - 2.43 tutoring sessions per student per week
SY 2023-2024	- 5 vendors - 36 LEAs - 131 schools - 5,244 students	- 219,809 tutoring sessions - 44.81 tutoring sessions per student - 2.88 tutoring sessions per student per week

Source: DESE Early Literacy Tutoring Studies (2023 & 2024), American Institutes for Research

Across the first two years of the study, AIR documented statistically significant student gains in literacy outcomes during the school year in which the students received the high-impact tutoring. Specifically, AIR documented the following impacts on student outcomes:

- The percentage of students performing below benchmark decreased from 84% to 59%

[1] Vendors went through a state-level pre-qualification process that required them to demonstrate that their tutoring model uses evidence-based practices for early literacy, as well as present evidence of: 1) capacity to serve students across Massachusetts; 2) demonstrated experience in successful implementation of tutoring in similar conditions; 3) demonstrated impact on student learning outcomes for similar sets of students. DESE approved 5 vendors that met the definition of "high-impact tutoring" as presented in Robinson, Carly D. and Susanna Loeb. (2021) High-Impact Tutoring: State of the Research and Priorities for Future Learning. (EdWorkingPaper: 21-384). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/qf76-rj21. High-impact tutoring is also high-dosage, meaning that a certain dosage of tutoring is required to reliably drive impact.

(SY22-23 and SY23-24).

- The percentage of students performing at or above benchmark increased from 16% to 41% (SY22-23 and SY23-24).
- SY2023-2024: on average, students' composite scores on DIBELS increased by 18 points (SD = 19).
- SY2023-2024: on average, students' composite scores on DIBELS increased by 19 points (SD = 21).
- Students with lower pretest scores showed stronger literacy gains than students with higher baseline scores (SY22-23 and SY23-24).

The AIR study demonstrated that DESE's Early Literacy Tutoring Program implements evidence-based tutoring practices with sufficient fidelity to achieve accelerated literacy growth for students.

Despite the growing evidence supporting the efficacy of tutoring, including the AIR study in Massachusetts specifically, practitioners still face important unanswered questions about high-dosage literacy tutoring (HDLT), including: ***Are these student gains in literacy outcomes sustained beyond the year in which students receive the high-dosage tutoring?*** BOOST aims to answer this question by providing HDLT to two cohorts of 1st graders (each cohort of the research study will include 21 schools and 420 students for a total of 840 students receiving HDLT) and measuring student literacy outcomes across the two years following treatment (2nd and 3rd grade). As discussed in the Evaluation section of this proposal, the intervention will also be measured in relation to a randomized control group to meet What Works Clearinghouse's evidence standards without reservation.

This study will also be able to address a critical issue raised by Nelson, Klingbeil, VanNorman and Parker (2021): “The bidirectional movement of students between tiers of support is fundamental to MTSS: however, much more is currently known about when to increase resources than when to decrease resources for students.” Is one year of HDLT enough to sustain results? Are there particular aspects of HDLT that correlate with sustained outcomes? By studying two student cohorts both during HDLT (1st grade) and after (grades 2 and 3), BOOST will be able to provide insights to these questions and inform broader field implementation of such an intensive intervention. This will support design, implementation and scaling of HDLT with cost-effective, evidence-based strategies that sustain student outcomes.

Intervention 2: Add Low-Dosage Literacy Tutoring in 2nd Grade

While evaluation of Intervention 1 will help address how to better design and implement high-dosage tutoring, we know from research that additional follow-up interventions are typically required to sustain skills gained in the intervention. Nelson et al (2018) found that 40% of students who successfully exited from a Tier 2 reading intervention regressed to an “at-risk” level when assessed at year-end even though they had performed above year-end benchmarks at the time that they exited the intervention. They suggest that educators consider developing “sustaining environments” (Bailey et al., 2018, 2020) with practices to protect against post-intervention regression. Nelson et al., drawing from Burns et al, 2008 and Stokes & Osnes, 1989, argue that “with specific academic skills such as early literacy, well-researched instructional practices including providing additional opportunities to respond, may help approximate the sustaining conditions necessary for skill maintenance.”

Therefore, in addition to identifying and mitigating factors for regression in the design and delivery of HDLT, BOOST will study the <u>addition of another, low-cost and highly scalable intervention: low-dosage literacy tutoring (LDLT) during the year following HDLT.</u> Adding this intervention is based on promising practices identified in research. In a 2021 randomized controlled trial, Nelson et al. examined spring literacy outcomes for early grade students who successfully exited an evidence-based tutoring program during the fall semester. Successful responders in K, 2 and 3 were randomly assigned by schools to receive either no follow-up support or access to a once-weekly oral practice session (lighter-dosage intervention). While results were mixed across grades and outcomes, treatment in K and 3rd grade corresponded to ≅15% increased likelihood of meeting year-end benchmarks on targeted literacy skills.

BOOST proposes providing both Intervention 1 (HDLT in grade 1) and Intervention 2 (HLT in grade 1 + LDLT in grade 2) for two additional cohorts of students (each cohort will include 21 schools and 420 students for a total of 840 students). These two treatment groups will be compared to a control group ("business as usual") of 21 schools and 420 students. We will measure student literacy outcomes during both intervention years (1st and 2nd grades) and the year following treatment (3rd grade). This component of the study will help determine the efficacy of a cost-effective intervention to sustain the impacts of grade 1 tutoring through grades 2 and 3, which can be scaled in other early literacy programs across the country. Building on the Nelson (2021) study, BOOST will add to the evidence base by studying a larger number of students across two cohorts and across multiple years.

Significance of the BOOST Innovation

By conducting a randomized controlled trial (RCT) of students that only receive Intervention 1 (HDLT in grade 1) and students who receive Intervention 1 plus Intervention 2 (HDLT in grade 1 and LDLT in grade 2), BOOST will be able to address critical issues in tutoring, including:

- <u>Short-term impact</u>: What is the impact of one year of high-dosage tutoring on students' early literacy achievement compared with business-as-usual instruction?

- Extended impact: At the end of second grade, how do literacy outcomes differ among students who received one year of tutoring, one year of tutoring with a light-touch continuation, or business-as-usual instruction?
- Persistence: To what extent do literacy impacts persist after tutoring ends for each group into third grade?
- Moderators: How do impacts vary by student characteristics (e.g., baseline literacy proficiency, grade) and by school context (e.g., urbanicity, charter status, staffing patterns)?
- Fidelity–outcome linkages: How is implementation fidelity associated with student reading achievement?
- Cost and cost-effectiveness: What are the total and marginal costs of delivering one year of tutoring versus one year of tutoring with a light-touch continuation, and what is the cost per unit of literacy gain?
- Scalability and replication: What implementation conditions and resource requirements are necessary to replicate the model across diverse Massachusetts school contexts?

Strategy to Scale

Unmet Demand for Broader Implementation

BOOST is an extension of larger statewide efforts to address a troubling trend in early grade literacy in Massachusetts. From 2005 to 2017, the National Assessment of Educational Progress (NAEP) showed no significant growth in Massachusetts 4th grade reading achievement. Since 2017, reading results in 4th grade have declined to below 2005 levels. (See Figure 2.) On the annual state assessment, the Massachusetts Comprehensive Assessment System English Language Arts (MCAS ELA), we see similar trends for 4th grade students over time as on the NAEP Reading. The percent of students meeting or exceeding expectations in ELA remained relatively flat since 2017 and steeply declined following the pandemic in 2019 (see Figure 3).

Figure 2: Massachusetts 4th grade NAEP Reading Scores 2005-2024

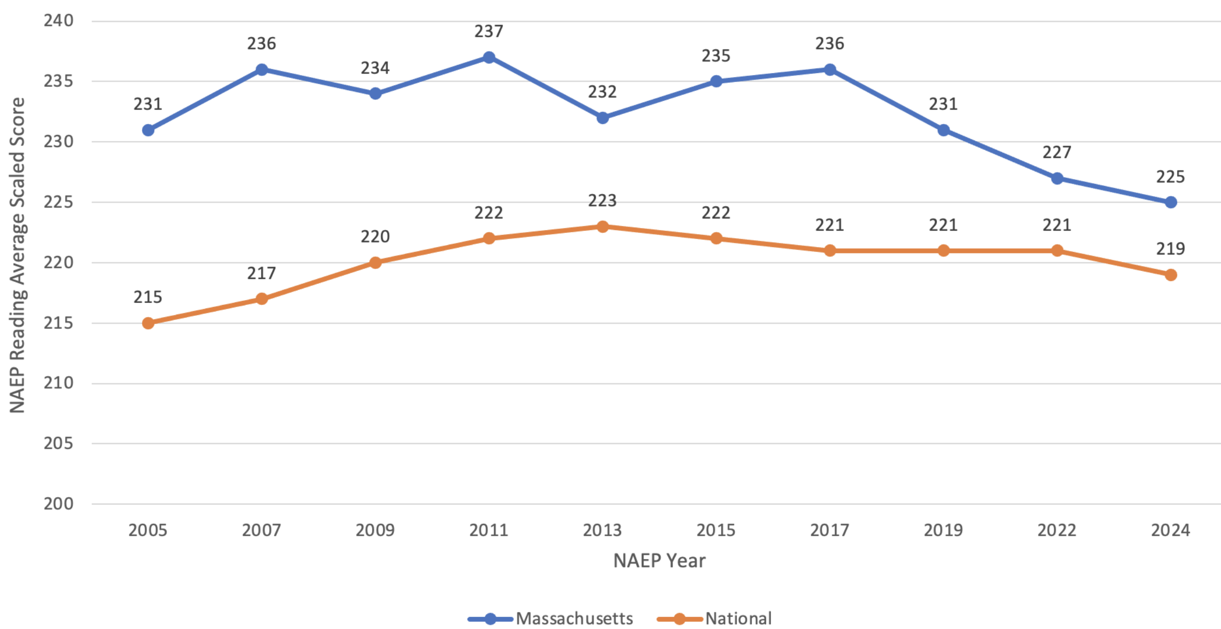
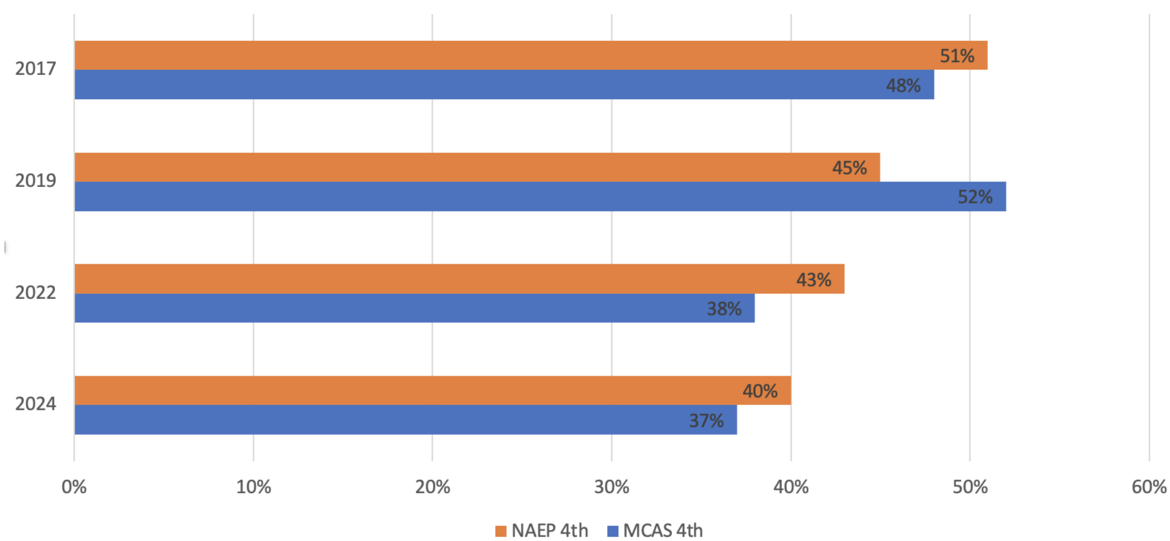


Figure 3: 4th grade NAEP Reading & MCAS ELA Proficiency, 2017-2024



Massachusetts has responded vigorously to these worrying trends first by developing a State Literacy Strategic Plan in 2019 and then implementing a series of programs designed to move the state forward systematically in addressing the issues. These efforts include establishing clear policies, practices and procedures to identify “at risk” students in relation to literacy and reading proficiency. State regulation 603 CMR 28.03(1)(f), which became effective on July 1, 2023, states that “each school district shall at least twice per year assess each student's reading ability and progress in literacy skills, from kindergarten through at least third grade, using a valid, developmentally appropriate screening instrument approved by the Department.” Based on the scores reported to DESE in SY23-24 (the most recent year for which data are available), the states estimates that 45% of Kindergarten students, 48% of 1st graders, 45% of 2nd graders, and 46% of 3rd graders were categorized as “at risk” at least two out of three times they were assessed, which represents a total of approximately 118,713 students across Massachusetts.

As part of its comprehensive efforts to address the needs of these at-risk students, DESE launched the Early Literacy Tutoring Program, which has provided high-impact, high dosage literacy tutoring (HDLT) to just under 16,000 high need[2] students in grades K-3 over the past five years and is slated to provide tutoring to 10,000 more high need students during SY2025-2026. This increasing scope includes expansion from 12 LEAs and 35 schools in SY20-21 to 66 LEAs and 325 schools in SY25-26, funded this year by a $25M state appropriation. Furthermore, demand for this tutoring dramatically outpaces those that DESE has been able to serve. For SY25-26, DESE received applications from 121 LEAs (114 traditional districts and 7 charter schools), representing approximately 473 schools and 27,267 students. Statewide, this represents 38% of all LEAs and 26% of all schools in Massachusetts requesting HDLT for their at-risk students.

Returning to the need data, Massachusetts currently has at least 100,000 students in

[2] As stated in “Significance,” DESE identifies “high need” for the statewide tutoring program (including the proposed BOOST program) as students who score below benchmark on DIBELS in the same year as tutoring is delivered.

grades K-3 who meet the definition of “high need” for early grade literacy tutoring, which is 10 times larger than those currently being served through the Early Grade Literacy Tutoring program. If funded, BOOST will add 4,100 students to those receiving tutoring. But that is not the main purpose of the grant. Rather, BOOST will identify, document and disseminate early grade tutoring strategies that are both cost-effective to implement and have the highest likelihood to sustain the impact of tutoring for the long term. This would reduce the significant costs associated with regression, including the need for additional HDLT in later years, making it possible to scale more cost-effective tutoring interventions in Massachusetts and nationwide.

Management Plan

Similar to other state leaders in the "science of reading" movement, Massachusetts has invested in spreading evidence-based early literacy practices. The proposed BOOST program will be part of Massachusetts’ comprehensive, research-driven efforts to improve early grade literacy described above. These efforts include: increasing the capacity of schools and teachers to use high-quality, evidence-based early literacy curricula, reliably identifying “high need” students for referral to early grade literacy interventions, and providing high-impact early grade literacy interventions, including high-dosage literacy tutoring (HDLT). (See the State Literacy Plan Executive Summary and relevant State Regulations in Appendix J.) Collectively, these initiatives provide a strong statewide foundation for Massachusetts to implement the proposed innovation with fidelity, on time and within budget.

BOOST will be implemented across five years, beginning with a planning period from January 2026 to July 2027. The pilot study will run from August 2027 to June 2028 with 590 high need 1st grade students receiving HDLT and 150 high need 2nd grade students receiving LDLT. The pilot will allow us to field test and refine LDLT as well as our existing Statewide

Tutoring Implementation Toolkit to support the full study implementation, which begins in SY27-28. As depicted in Figure 4, below, we will use a cohort model to implement the full study. Cohorts A and B will each include 63 schools split into 3 groups: a control group of 21 schools and 420 students who do not receive any interventions; treatment group 1 of 21 schools and 420 students who receive HDLT in grade 1, and; treatment group 2 of 21 schools and 420 students who receive both HDLT in grade 1 and LDLT in grade 2. We will track student outcomes for all three groups through 3rd grade to measure persistence of impact for both HDLT and LDLT.

Figure 4: Pilot & Cohort Implementation

	SY26-27	**SY27-28**	**SY28-29**	**SY29-30**	**SY30-31**
	Pilot Study	**Study**	**Study**	**Study**	**Research Close-out & Dissemination**
Control (No Tutoring)		Cohort A Control Group 21 schools 420 students (1st grade)	Cohort A Control Group 21 schools 420 students (2nd grade)	Cohort A Control Group 21 schools 420 students (3rd grade)	
			Cohort B Control Group 21 schools 420 students (1st grade)	Cohort B Control Group 21 schools 420 students (2nd grade)	Cohort B Control Group 21 schools 420 students (3rd grade)
1st Grade HDLT	Pilot HDLT 590 students	Cohort A Treatment Groups 1 & 2 42 schools 840 students	Cohort B Treatment Groups 1 & 2 42 schools 840 students (1st grade)		
2nd grade LDLT	Pilot LDLT 150 students		Cohort A Treatment Group 2 21 schools 420 students (2nd grade)	Cohort B Treatment Group 2 21 schools 420 students (2nd grade)	
Continued tracking of treatment groups			Cohort A Treatment Group 1 21 schools 420 students (2nd grade)	Cohort A Treatment Groups 1 & 2 42 schools 840 students (3rd grade)	

				Cohort B Treatment 1 21 schools 420 students (2nd grade)	Cohort B Treatment 1 & 2 42 schools 840 students (3rd grade)

BOOST will be implemented as part of the existing Early Grade Literacy Tutoring Program, which has been operating since 2021 and has a strong research base demonstrating effective implementation of a statewide tutoring program (see “Significance”). BOOST will be led by a full-time **BOOST Project Director** who will oversee all aspects of implementation and work directly with **Local Education Agency (LEA) and School Coordinators** to ensure tutoring is implemented with fidelity. This includes revising and expanding Massachusetts’ existing Statewide Tutoring Implementation Toolkit, which has been developed with input from educators over the past five years and includes guidance for working with tutoring vendors, engaging LEAs, launching and monitoring tutoring, and data collection policies, procedures, tools and templates. The Project Director will also work directly with the **Evaluators** to carry out all aspects of the randomized control trial (RCT), review formative implementation data to inform continuous improvement, and post reports in a timely manner to the DESE website.

The Project Director will report directly to [redacted]. Other DESE **Leadership Team Members** on the project include [redacted], [redacted], psychometrician and measurement expert; [redacted], who will support coordination with DESE’s budget, HR, and communications teams to ensure adequate resourcing for smooth implementation of BOOST; [redacted], who will serve as evaluation plan lead, and [redacted], who will coordinate relationships with participating LEAs, schools, and tutoring providers and oversee provision of technical assistance to schools.

Under the direction of [redacted], this team will ensure that state resources and infrastructure are mobilized to support the project. Please see Appendix B for resumes of these key personnel.

We have developed a comprehensive Management Plan that is aligned with the goals, objectives and outcome measures included in "Project Design" and includes specific activities with start and end dates as well as the responsible party for each activity. Please see Appendix J for the detailed Management Plan.

Partners & Support

BOOST builds upon existing, long-standing and effective partnerships between DESE and state-approved providers of high-impact literacy tutoring. DESE has been working with most of these providers for over 5 years and has deep relationships and existing coordination systems to lean on for this expansion. We also propose to extend an existing relationship with Johns Hopkins University, our research partner currently evaluating the statewide tutoring program. Our school districts have advocated energetically for additional tutoring resources, and the Massachusetts Association of School Superintendents (M.A.S.S.), representing the state's district leaders, supports BOOST. We designed BOOST with input from our partners to ensure feasible implementation of the proposed project. Appendix C includes letters of support from all key partners, documenting past successful partnerships and commitments to the success of BOOST.

Efficient Delivery at Scale

Massachusetts is in a strong position to scale high-impact tutoring both efficiently and effectively during the proposed grant period. As discussed in "Strategy to Scale," DESE has built a statewide infrastructure to implement evidence-based literacy instruction, including development of a statewide literacy action plan in 2019, implementing a comprehensive statewide literacy development program (2020 to the present), instituting teacher training policies so that educators are ready to implement evidence-based literacy instruction in all classrooms, and implementing policies and regulations that support identification of high need students in grades K-3 through early grade literacy assessments.

Furthermore, BOOST builds on the documented success of the Early Literacy Tutoring Program, which has delivered high-impact tutoring in the form of HDLT to approximately 16,000 students since 2021 and is expanding to 10,000 more students in SY25-26. As discussed in "Significance," this project has demonstrated effectiveness as it scales. For example, when the project increased from 67 schools and 2,907 students in SY22-23 to 131 schools and 5,244 students in SY23-24, the dosage of tutoring increased from 37.93 tutoring sessions per student to 44.81 sessions and the average number of tutoring sessions per student per week also increased from 2.43 to 2.88. These increases were based on efficiencies in implementation as the per-student costs paid to vendors did not increase across the two school years.

BOOST will leverage the statewide infrastructures that support evidence-based literacy instruction and high-impact early grade literacy tutoring to deliver project services efficiently at scale and maintain effectiveness. In particular, we will refine and expand our Statewide Tutoring Implementation Toolkit, which has been critical to efficient and successful scaling of HDLT. In Year 5 of the grant, we will develop a replicable version of this toolkit to share with other entities working to implement high-impact tutoring at scale.

Dissemination Plans

In 2021, DESE built a comprehensive website dedicated to "Mass Literacy"—empowering educators with evidence-based practices for literacy that all students need. The website includes sections addressing components of the core literacy block, skills for early reading, students experiencing reading difficulties, leading MTSS, early literacy in educator preparation, and resources for families and communities. The website includes information in multiple formats, including interactive infographics and layered presentations of evidence-based practices that allow users to dig deeper into specific areas of interest. The language and format are user-friendly and accessible to a wide range of audiences, including families, communities, educators, policy makers, and researchers.

We will build out a section of the DESE website to share information, resources and materials related to BOOST, including all research reports and key implementation tools. We will post materials in a timely manner, measure scope of impact through download data on users (numbers, locations, etc.), and user satisfaction with resources through point-of-use surveys (see "Management Plan" in Appendix J). See "Evaluation" for more on dissemination of research.

Project Design

Logic Model

Please see Appendix G.

SMART Goals

We have six project goals, each with objectives and corresponding outcome measures. Please see Appendix J for the full "Management Plan," which directly aligns with these SMART goals and includes activities associated with achieving them.

Goal 1 - Expand access to high-impact tutoring for high need students (defined as students scoring below benchmark on DIBELS in the same year as receiving HDLT)

- Ob 1.1 - Implement BOOST with 740 high needs students in Pilot Study
  - OM 1.1.a - 100% of students in pilot study meet definition of high need student
- Ob1.2: Implement BOOST with 1,260 high need students in Cohort A schools, consisting

of 3 groups: Control Group - 21 schools and 420 students in control (“business as usual”); Treatment Group 1 - 21 schools deliver HDLT in 1st grade only to 420 students; Treatment Group 2 - 21 schools deliver HDLT in 1st grade and follow-up LDLT in 2nd grade to 420 students

    - OM 1.2.a - 100% of students in Cohort A meet definition of high need student

- Ob 1.3 - Implement BOOST with 1,260 high need students in Cohort B schools, consisting of 3 groups: Control Group - 21 schools and 420 students in control (“business as usual”); Treatment Group 1 - 21 schools deliver HDLT in 1st grade only to 420 students; Treatment Group 2 - 21 schools deliver HDLT in 1st grade and follow-up LDLT in 2nd grade to 420 students
    - OM 1.3.a - 100% of students in Cohort B meet definition of high need student

Goal 2 - Measure impact of 1 year of HDLT vs. 1 year of HDLT+1 year of LDLT continuation through randomized controlled trial (RCT)

- Ob 2.1 - Develop and pilot evaluation instruments
    - OM 2.1.a - 100% of evaluation instruments are field-tested prior to use in full evaluation study
- Ob 2.2 - Implement evaluation study
    - OM 2.2.a - Three formative evaluation reports include actionable data to inform ongoing implementation of BOOST
    - OM 2.2.b - One summative evaluation report includes actionable data to inform broader field implementation of high-impact tutoring
- Ob 2.3 - Accelerate and maintain literacy growth for treatment students
    - OM 2.3.a - >60% of students receiving HDLT in 1st grade achieve accelerated

growth on DIBELS during the treatment year

  - OM 2.3.b - >60% of students receiving HDLT in 1st grade maintain or improve growth status on DIBELS in 2nd and 3rd grades
  - OM 2.3.c - >60% of students receiving HDLT in 1st grade and LDLT in 2nd grade accelerate and maintain literacy growth at higher rates than those who only receive HDLT in 1st grade

Goal 3 - Identify dosage, duration, and contextual factors driving sustained literacy gains

- Ob 3.1 - Examine persistence of tutoring impacts after direct services conclude
  - OM 3.1.a - Three persistence of impacts reports identify dosage and duration factors driving sustained literacy gains
- Ob 3.2 - Examine contextual factors associated with sustained literacy gains.
  - OM 3.2.a - Three contextual factor reports will identify contextual factors associated with sustained literacy gains.

Goal 4 - Optimize statewide tutoring investments through sustainable, scalable delivery models

- Ob 4.1 - Identify indirect costs for implementing HDLT and LDLT tutoring
  - OM 4.1.a - Annually, >80% of participating schools and LEAs deliver completed work logs and documentation of un-budgeted costs
- Ob 4.2 - Conduct comprehensive cost-analysis for HDLT and LDLT tutoring
  - OM 4.2.a - Three formative cost analysis reports measure cost-effectiveness of BOOST and identify best practices for improving cost-effectiveness that are generalizable to the field.
  - OM 4.2.b - One summative cost analysis report measures cost-effectiveness of

BOOST and identifies best practices for improving cost-effectiveness that are generalizable to the field.

Goal 5 - Develop evidence-based toolkit to support implementation of high-impact tutoring at scale

- Ob 5.1 - Refine and expand existing statewide implementation toolkit through pilot implementation
    - OM 5.1.a - >75% of school coordinators in pilot will report consistent use of statewide implementation toolkit
    - OM 5.1.b - Revised and expanded toolkit will incorporate feedback from pilot implementation
- Ob 5.2 - Refine statewide implementation toolkit through study implementation
    - OM 5.2.a - >75% of school coordinators in pilot will report consistent use of statewide implementation toolkit
    - OM 5.2.b - Revised and expanded toolkit will incorporate feedback from study implementation
- Ob 5.3 - Develop and disseminate replicable statewide tutoring implementation toolkit for use beyond Massachusetts
    - 0M 5.3.a - Replicable toolkit incorporates findings from summative assessment data

Goal 6 - Disseminate resources and findings to support scaling of high-impact tutoring beyond Massachusetts

- Ob 6.1 - Disseminate research findings in formats comprehensible to both practitioners and the general public

    - OM 6.1.a - >75% of digital users who complete point-of-use surveys will report materials to be comprehensible and useful
- Ob 6.2 - Promote replication and transparency of RCT study
    - 0M 6.2.a - 100% of publicly releasable materials, including codebooks, fidelity instruments, and de-identified analytic code, are archived on the Open Science Framework

Project Evaluation

The Center for Research and Reform in Education (CRRE) will conduct an independent evaluation of the impact and implementation of the BOOST high-impact tutoring initiative in Massachusetts on elementary reading achievement using a cluster randomized control trial (RCT) to meet What Works Clearinghouse (WWC) standards without reservations (What Works Clearinghouse, 2022). CRRE is a research center in the School of Education at Johns Hopkins University that has conducted hundreds of evaluation studies and has the stated goal of improving the quality of education for students through high-quality research and evaluation studies and the dissemination of evidence-based research. DESE has successfully partnered with CCRE to evaluate projects in the past, including a current study of the Early Grade Literacy Program. This existing collaboration will help streamline the proposed BOOST evaluation.

The proposed evaluation will test the short- and longer-term impacts of two tutoring approaches—(1) a **one-year, high-dosage literacy tutoring (HDLT) model** delivered in first grade, and (2) the same HDLT first-grade tutoring model **paired with low-dosage literacy tutoring (LDLT) continuation in second grade**—against a **business-as-usual (BAU)** condition representing the literacy supports and interventions schools typically provide in Massachusetts. The study will also examine implementation fidelity, mechanisms of impact, and cost-effectiveness to provide actionable guidance for replication and scaling. The project focuses on the following research questions.

1. Short-term impact: What is the impact of one year of HDLT on students' early literacy achievement at the end of that year, compared with business-as-usual instruction?
2. Extended impact: At the end of second grade, how do literacy outcomes differ among students who received one year of HDLT, one year of HDLT + one year of LDLT continuation, or business-as-usual instruction?
3. Persistence: To what extent do literacy impacts persist after tutoring ends for each group into third grade?
4. Moderators: How do impacts vary by student characteristics (e.g., baseline literacy proficiency, grade, socioeconomic status, race/ethnicity, English learner status) and by school context (e.g., urbanicity, charter status, staffing patterns)?
5. Fidelity–outcome linkages: How is implementation fidelity associated with student reading achievement?
6. Cost and cost-effectiveness: What are the total and marginal costs of delivering one year of tutoring versus one year of tutoring with a light-touch continuation, and what is the cost per unit of literacy gain?
7. Scalability and replication: What implementation conditions and resource requirements are necessary to replicate the model across diverse Massachusetts school contexts?

These research questions align with the project's objectives and will be addressed using data collected from schools across Massachusetts.

Methods of Evaluation Will Produce WWC-Standard Evidence

Study Design and Random Assignment

The proposed evaluation will employ a cluster randomized controlled trial (RCT) at the school level to assess the short-term, extended, and persistent impacts of one year of HDLT versus one year of HDLT plus LDLT continuation versus business-as-usual on early literacy achievement. This design, if well executed, will meet *What Works Clearinghouse* (WWC) standards without reservations (What Works Clearinghouse, 2022). Participating elementary schools will be randomly assigned within districts to one of three conditions:

1. <u>Control (Business-as-Usual - BAU)</u>. Schools will continue implementing their existing literacy instruction and supports within a school-based multi-tiered system of supports (MTSS) framework. This includes Tier 1 core literacy instruction and Tier 2 small-group interventions.
2. <u>One-Year High Dosage Literacy Tutoring (HDLT)</u>. Students will receive HDLT for one academic year in first grade, after which they return to BAU instruction in second grade.
3. <u>One-Year HDLT + Low Dosage Literacy Tutoring (LDLT) Continuation</u>. Students will receive HDLT in first grade and LDLT in second grade, designed to reinforce foundational skills and maintain literacy growth.

Randomization will occur at the school level, the policy-relevant unit for tutoring implementation and decision-making, to prevent contamination that could occur if students within the same school were assigned to different conditions. Blocking within districts will ensure balance on key characteristics such as enrollment size, baseline literacy performance, and locale. Schools will be randomized to a condition and then they will stay in that condition, with the second cohort adding more students to each school's participating group.

The study will include 63 schools (21 per arm), each enrolling approximately 20 high

need[3] first-grade students per cohort across two cohorts (Cohort A beginning in SY 2027–28 and Cohort B in SY 2028–29). This yields a total analytic sample of 2,520 students overall. Schools will represent a cross-section of Massachusetts's urban, suburban, and rural contexts as well as traditional public and charter schools, supporting both internal validity and generalizability.
The evaluation timeline spans five years. Planning and refinement of evaluation instruments will occur during SY 2026–27, followed by two years of implementation and data collection (SY 2027–28 and SY 2028–29). Follow-up data collection will continue through SY 2029–30 to assess persistence for Cohort A. Final analyses and reporting will be completed in SY 2030–31.

For the impact analyses, the minimum detectable effect size, or MDES (the smallest true effect that can be detected for a specified level of power and significance level for any given sample size) is estimated for reading achievement impacts for the comparison of treatment versus control in this cluster randomized trial at different timepoints. Using PowerUp! (Dong et al., 2013), we carried out power analyses using estimated values of ICC and R^2 from Hedges and Hedberg (2007), an alpha level of .05, a two-tailed test, power of .80, an starting average cluster size of 40, no school-level, and 15% student-level attrition after each year. The MDES for the overall impact analysis for the impact of 1 year of HDLT at the end of the first year is +0.148. The MDES for the impact of HDLT (1 year) + LDLT (1 year) compared to control at the end of the second year is +0.176, increasing to +0.205 by the end of the third year for impacts sustained after 1 year of tutoring. Given the prior research on high-dosage tutoring alone, we believe the study is adequately powered to detect effects (Cortes et. al, 2025; Kraft et. al, 2024).

By embedding random assignment within authentic district structures and maintaining rigorous design standards for implementation, measurement, and analysis, the BOOST

[3] "High need" refers to students achieving below benchmark on DIBELS in the same year that they receive HDLT.

evaluation will produce internally valid, policy-relevant evidence on the effectiveness of high-impact tutoring that meets WWC standards without reservations.

Outcomes and Measures

The evaluation will assess student literacy achievement using a combination of standardized assessments and detailed implementation and instructional data. Primary outcomes will focus on early reading proficiency measured through the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) at the end of the first and second years of tutoring. These measures provide sensitive indicators of phonemic awareness, decoding, and fluency that have been used in prior large-scale tutoring evaluations. Longer-term impacts will be measured using Massachusetts' state assessment, the MCAS English Language Arts test, administered in the spring after the follow-up year.

In addition to student outcomes, the evaluation will collect rich, session-level implementation data to capture the nature, dosage, and quality of tutoring delivered. This information will be drawn systematically from tutoring logs, online platform data, and observation protocols, enabling detailed analysis of how fidelity and dosage relate to student outcomes. These tools have been used in prior implementations and provide valid and reliable data on implementation. To contextualize these findings, the study will also document the broader literacy instruction and intervention landscape available to students across all conditions, including the control group, using an adaptation of a previously validated survey of literacy supports (Jacob et al., 2016). This survey instrument will provide a comprehensive picture of the instructional resources, interventions, and curricula in use across participating schools, ensuring that the evaluation accurately characterizes what "business-as-usual" entails in Massachusetts' high-impact literacy tutoring context.

Together, these measures will permit rigorous estimation of both the immediate and sustained effects of one and two years of tutoring while providing insight into the instructional conditions and implementation practices that influence those impacts.

Analysis Plan

The evaluation will use an intent-to-treat (ITT) framework to estimate the causal effects of one year of HDLT and one year of HDLT plus one year of LDLT relative to business-as-usual instruction. Analyses will account for the nested structure of the data, with students nested within schools, using hierarchical linear models that include school-level random effects and baseline covariates to improve precision and reduce residual bias.

The analytic approach will estimate three sets of contrasts corresponding to the study's research questions: (1) the impact of one year of HDLT relative to the control group at the end of the first implementation year, (2) the relative and cumulative effects of one year of HDLT and one year of HDLT + one year of LDLT at the end of the second year, and (3) the persistence of impacts one and two years after tutoring ends, measured through DIBELS and state assessment outcomes.

All analyses will follow pre-specified protocols to ensure transparency and replicability. The evaluation will be pre-registered on the Open Science Framework prior to baseline data collection. Models will include baseline literacy scores, grade-level, and demographic characteristics as covariates.

To understand how impacts vary across contexts and student groups, the evaluation will estimate moderation effects by student characteristics (e.g., baseline literacy proficiency, socioeconomic status, race/ethnicity, and English learner status) and by school-level factors such as support staff availability and resource levels. Exploratory analyses will examine the relationship between tutoring dosage and student achievement.

Threats to Validity and Mitigation

The evaluation is designed to produce internally valid and policy-relevant evidence under real-world conditions. Random assignment at the school level prevents contamination across students and eliminates selection bias, and external vendor implementation prevents contamination across schools within the district.

One design feature that reduces the risk of cluster-level attrition is the school-level randomization strategy. Because districts agree to participate with the assurance that all will receive tutoring services in some capacity, they face no risk of being excluded from the intervention entirely. This design encourages sustained district engagement across all study years. Furthermore, even schools assigned to the control condition will receive tutoring services in unmeasured grades (e.g., upper elementary grades not included in the study sample), maintaining district goodwill and minimizing dropout or withdrawal from the evaluation. The use of external vendors, rather than school-employed tutors, further limits the potential for cross-school contamination within districts while ensuring consistent implementation fidelity. Student-level loss of sample will be minimized through DESE's existing data systems that routinely collect these data statewide via tutoring providers. Student mobility will be tracked to maintain analytic integrity and expected attrition rates fall below the What Works Clearinghouse thresholds for low and balanced attrition.

Variation in implementation will be monitored through reviews of tutoring dosage provided to ensure students are receiving the intended dosage of tutoring (HDLT or LDLT). DESE has years of experience building and refining systems for collecting and validating tutoring dosage logs from providers. These data will also support interpretation of results and link implementation quality to student outcomes.

Finally, comprehensive contextual data on school resources and literacy supports across all conditions will help ensure that findings are both credible and generalizable. Together, these procedures will allow the study to meet WWC standards without reservations.

Evaluation Will Provide Guidance for Replication and Broader Implementation

Replication Focus

The BOOST evaluation is designed to generate practical, transferable evidence on how high-impact literacy tutoring can be implemented and sustained at scale within Massachusetts's MTSS framework. By including a diverse mix of urban, suburban, and rural schools across the state, the study tests BOOST under authentic implementation conditions, ensuring findings reflect both effectiveness and feasibility in real-world contexts.

A central replication feature of the study is the comparison of a full-year of HDLT to a full-year of HDLT + a year of LDLT continuation. This design directly addresses questions of sustainability and scalability—whether schools can maintain or extend literacy gains with reduced intensity and cost once initial implementation capacity is built. Evidence from this contrast will inform states and districts seeking to expand tutoring while managing resource constraints, offering concrete guidance on the conditions under which an LDLT model can sustain early literacy achievement.

The inclusion of two cohorts allows for replication of results across successive implementation cycles, providing additional insight into the consistency and generalizability of impacts. Together, these design elements position BOOST to produce evidence not only of impact but of how tutoring can be adapted and sustained as a long-term, scalable literacy strategy.

Deliverables and Knowledge Products

The evaluation will produce a set of materials designed to guide replication and scaling in Massachusetts and beyond. Core deliverables will include a refined and expanded Statewide Tutoring Implementation Toolkit that translates evaluation findings into actionable guidance for state and district leaders, documenting the practices, supports, and conditions associated with strong tutoring outcomes. In addition, the evaluation team will develop replication-ready materials such as fidelity rubrics and data collection templates to support future adopters. Annual interim briefs will communicate emerging lessons to state and local partners, while the final technical report will provide a comprehensive summary of impact, implementation, and cost findings. All products will be written in accessible language and shared through public channels, ensuring that the results contribute not only to Massachusetts' literacy strategy but also to the national evidence base on effective, scalable tutoring interventions. Please see "Management Plan" for specific reports that will be issued throughout the grant period.

Measuring Fidelity of Implementation Plan

Core Components and Active Ingredients

The evaluation will document the essential components of the tutoring model that drive improvements in early literacy outcomes. These core elements include the structure and frequency of tutoring sessions, the use of high-quality, evidence-based instructional materials, systematic tutor training and coaching, and the emphasis on consistent relationships between tutors and students. Together, these elements constitute the "active ingredients" hypothesized to produce gains in reading achievement. All approved tutoring vendors in Massachusetts provide services that incorporate these core elements and design features.

Fidelity Measures and Thresholds

Fidelity of implementation will be measured through multiple sources. Tutoring logs and session-level data will provide information on the frequency and duration of sessions, allowing the evaluation to document the extent to which students receive the intended dosage. Thresholds for acceptable implementation will be pre-specified and aligned with prior tutoring research. Schools will be considered to have met dosage fidelity when students receive at least 80 percent of scheduled sessions. These measures will allow the evaluation to determine whether impacts were achieved under conditions of strong implementation and to interpret variation in outcomes in relation to fidelity across schools and cohorts.

Data Sources for Fidelity

The evaluation will draw on multiple complementary data sources to capture a comprehensive picture of implementation. Tutoring logs and digital platform records will provide information on dosage, attendance, and session length. Surveys of relevant stakeholders including administrators, teachers, and tutors will allow us to better understand the degree to which the identified core components are implemented and their perceived role in the overall program impacts. Instructional materials will also be examined against required program elements.

Analytic Use

Fidelity data will be integrated directly into the impact analyses to examine how variation in implementation relates to student outcomes. Measures of dosage will be included as moderators and covariates in exploratory models to assess whether stronger implementation is associated with larger literacy gains. Fidelity data will also inform continuous improvement throughout the project. Interim analyses will identify schools or tutors that fall below fidelity thresholds, allowing implementation partners to provide targeted support. By linking fidelity to impact and feeding results back into program refinement, the evaluation will generate both rigorous causal evidence and actionable guidance for replication. Survey data will also provide additional information that will help to better contextualize and interpret findings.

Design Will Generate Information to Guide Replication and Scaling

Scalability and Cost-Effectiveness

The evaluation is designed to generate actionable information about the resources required to implement and sustain two models of high-impact tutoring at scale (1 year of HDLT and 1 year of HDLT+1 year of LDLT continuation). A comprehensive cost analysis will estimate both direct and indirect costs, including personnel, training, materials, technology, and administrative support. The analysis will identify hidden or variable costs, such as tutor recruitment and scheduling demands, that influence scalability in real-world conditions (Kohlmoos & Steinberg, 2025). Cost-effectiveness analyses will pair these estimates with measured impacts to calculate the cost per student served and the cost per 0.10 standard deviation improvement in literacy outcomes. These metrics will allow state and district leaders to compare the efficacy of the two tutoring models to each other and to alternative literacy interventions. By integrating cost data with fidelity and impact findings, the evaluation will provide a full picture of the resources, supports, and organizational conditions needed to scale the model equitably and sustainably across different types of school settings.

Persistence and Sustainability

The evaluation will examine the persistence of tutoring impacts after direct services conclude including both 1 year of HDLT and 1 year of HDLT + one year of LDLT continuation, providing critical evidence on the durability of early literacy gains. Follow-up analyses one and two years after tutoring will measure whether students maintain or continue to build on their reading achievement relative to control peers for each of the variations tested. These analyses will use policy-relevant district and state assessment data to assess long-term outcomes and inform policy decisions about the optimal duration and sequencing of tutoring supports.

Generalizability

The evaluation will assess the generalizability of findings by examining how participating schools compare to Massachusetts schools statewide on key demographic and contextual characteristics. Baseline analyses will document differences in enrollment size, student demographics, urbanicity, and instructional resources to clarify the populations and conditions to which results can be extended. The inclusion of various school types, spanning rural, suburban, and urban contexts, ensures that the findings reflect real variation in implementation environments. Analyses will test whether program impacts differ across these contexts, providing evidence on where and under what conditions the tutoring interventions are most effective. By combining rigorous impact estimates with detailed contextual data, the evaluation will produce evidence that is both credible and useful for policymakers and practitioners considering large-scale adoption.

Dissemination Plan

In addition to the dissemination work described in "Strategy to Scale," findings from the evaluation will be shared regularly with national, state, and district audiences, and especially schools, educators, and parents to ensure timely application of results and to empower families with information to advocate for services for their children. Annual interim briefs will summarize emerging implementation and impact findings for Massachusetts education leaders and participating schools. At project completion, the evaluation team will produce a comprehensive technical report detailing study design, analytic methods, impact estimates, and cost-effectiveness results.

To promote replication and transparency, all publicly releasable materials, including codebooks, fidelity instruments, and de-identified analytic code, will be archived on the Open Science Framework. A concise Implementation Playbook and summary briefs will translate key lessons for policymakers, school leaders, and practitioners, ensuring that the study's insights inform continued improvement and expansion of high-dosage tutoring.

Priorities

Absolute Priority - Promoting Evidence-based Literacy

BOOST focuses on a critical aspect of advancing evidence-based literacy: high-dosage tutoring. All vendors who deliver early grade literacy tutoring through BOOST will be pre-qualified by DESE to ensure that they meet the definition of "high-impact tutoring" as presented in Robinson, et al. (2021), High-Impact Tutoring: State of the Research and Priorities for Future Learning. Furthermore, as described in the Strategy to Scale section of this proposal, BOOST is embedded within a statewide approach to implementing evidence-based literacy instruction at scale. See the Evidence Form for the research base referenced in development of our proposed innovation.

Competitive Preference Priority 1

BOOST will be a statewide project, led by the State Education Agency, that builds on the statewide infrastructure for implementing early grade literacy that Massachusetts has established over the past five years (see "Strategy to Scale" for details).

Competitive Preference Priority 2

BOOST will deliver high-impact tutoring through innovative delivery models—HDLT and LDLT—in order to accelerate and maintain academic progress. The RCT study will help identify dosage, duration, and contextual factors for high-impact tutoring that drive sustained literacy gains at scale.