

A. Significance

Core Innovation: Generating the Evidence States Most Need

Louisiana’s Education Innovation and Research (EIR) Expansion project addresses three evidence gaps in high-impact tutoring (HIT):

1. Can an evidence-based virtual tutoring model deliver consistent results across varied locales (urban, suburban, town, rural) statewide?
2. What implementation characteristics sustain program quality when scaling from single-district success to statewide deployment under varied contexts?
3. Does extended literacy tutoring dosage justify added investment, and do gains persist once tutoring ends?

By rigorously testing these questions through a five-year randomized controlled trial (RCT) led by Johns Hopkins University’s Center for Research and Reform in Education (JHU-CRRE), Louisiana will produce actionable evidence to guide sustainable literacy policy nationwide.

Eligibility Foundation: Evidence-Based Literacy High-Impact Tutoring Program Grounded in the Science of Reading

Air Reading, the implementation partner for this project, delivers structured-literacy high-impact virtual literacy tutoring supported by strong evidence from a well-designed RCT demonstrating statistically significant literacy gains.

The Science of Reading (SOR)-aligned, skill-based program provides explicit, systematic, and intentional instruction in phonemic awareness, phonics, fluency, vocabulary, and comprehension. The high-impact tutoring program embodies small-group instruction (two–three students per tutor), consistent tutor assignment throughout the intervention period, four sessions

per week with 30 minutes each session, explicit instruction, ongoing progress monitoring that enables instructional adjustments, and sustained duration (20–24 weeks annually). The virtual model aligns tutoring content directly with Tier 1 core instruction through the Louisiana Student Standards for ELA and state-adopted curriculum (Guidebooks), ensuring coherence between intervention and classroom learning.

The Air Reading model's evidence base is robust: Air Reading delivers Science of Reading-aligned structured literacy through small-group virtual tutoring (1:2-1:3 ratios, four sessions weekly, 20-24 weeks). A peer-reviewed, What Works Clearinghouse (WWC)-standard RCT by JHU-CRRE demonstrated effectiveness in a high-need population (n=381, Grades 1-6, Texas): +0.12 SD overall gains and +0.20 SD in Grade 1, this project's focal grade ([REDACTED]). The study met WWC Group Design Standards without Reservations, and Air Reading appears on Evidence for ESSA as a Strong-Evidence program. An additional 2024-25 Louisiana RCT (n=170) using state assessments (DIBELS, LEAP) showed stronger effects with full-year implementation (+0.31 SD, $p < .05$), demonstrating both Louisiana-context effectiveness and the value of extended dosage ([REDACTED]).

Research Innovation 1: Cross-Context Replication at State Scale

Modification to serve different populations and settings. Two prior RCTs demonstrated Air Reading's effectiveness in homogeneous, single-district contexts: rural Texas (six schools, one semester) and suburban Louisiana (five schools, full year). No study has tested whether these effects replicate when Air Reading is deployed simultaneously across varied contexts at state scale. Louisiana will implement the model in 90 schools spanning urban, suburban, town, and rural locales (~4,500 students), including high- and lower-resourced schools, charters, and traditional districts, to answer:

- *Does the model maintain effectiveness under varied contexts at state scale?*
- *Under what conditions does it work best?*
- *What implementation characteristics (curriculum, tutor training, scheduling structures, instructional coherence, technical infrastructure, and school coordination) sustain instructional fidelity and program quality when scaling statewide?*

Establishing causal evidence requires controlled implementation contexts that maximize internal validity. Louisiana strategically varies context while implementing at scale to maximize external validity. The state’s varying locales, including rural parishes such as Concordia, Union, and Iberia, urban centers like East Baton Rouge and Jefferson, and charter schools such as Young Audiences Charter School, create an ideal testing ground for scalability. Results will identify the implementation conditions that predict the strongest effects and inform targeted technical assistance and resource allocation for sustainable statewide scaling.

Research Innovation 2: Testing Extended Dosage and Persistence

Extension of an evidence-based component. Prior research proves Air Reading’s one-year effectiveness, yet no RCT has compared one versus two consecutive years of tutoring or measured persistence beyond the intervention year. States invest millions in tutoring without knowing whether a longer duration yields sustained benefits.

Louisiana’s three-arm cluster RCT directly tests this policy question. Ninety schools (~4,500 students in Grades 1–2 below proficiency) will be randomly assigned to (1) business-as-usual, (2) one year of tutoring, or (3) two years of tutoring. Outcomes will be measured at year one, year two, one year after tutoring ends, and two years after tutoring ends. This design allows estimation of immediate, cumulative, and persistent effects and determines the cost per sustained literacy gain.

Most recent tutoring studies stop at post-test; Louisiana will track students through 2030, providing rare longitudinal evidence on durability. Findings will help policymakers decide whether to fund single-year or multi-year models and quantify the returns on extended investment. The evaluation is powered (MDES ~ 0.11–0.13) to detect educationally meaningful effects and meets WWC Standards without Reservations.

Implementation Innovation 1: Systemic Integration with Statewide Literacy Infrastructure

Strategic implementation within Louisiana's aligned state system. Most tutoring research evaluates standalone programs operating outside state systems. This project tests whether embedding HIT within an aligned SoR infrastructure strengthens both impact and sustainability.

Air Reading tutoring aligns directly with Louisiana’s High-Quality Instructional Materials (HQIM), uses the state’s universal literacy screener (DIBELS) for progress monitoring, and connects tutor professional development to existing educator learning networks. This design mirrors how states would realistically implement tutoring at scale.

The evaluation will test whether this systemic coherence between Tier 1 instruction, intervention, and professional learning produces stronger student outcomes and longer-term persistence than traditional “add-on” tutoring models. Findings will show whether integrating tutoring into a state’s literacy infrastructure amplifies impact, or whether stand-alone approaches are equally effective, providing states clear evidence on the most sustainable way to scale proven tutoring interventions.

Implementation Innovation 2: Workforce Expansion for Sustainable Scaling

Strategic workforce approach addressing capacity, cost, and teacher shortages. Louisiana will evaluate whether college students from various majors can deliver Air Reading effectively alongside professional tutors. College students receive identical training and supervision but are

compensated through work-study, service-learning credit, and stipends—potentially creating more abundant, affordable capacity, particularly valuable in rural areas. This approach also provides educational exposure that may spark teaching career interest: recent research shows tutoring experience increases teaching applications by 12 percentage points, with stronger effects for non-education majors (+14 points), suggesting potential dual benefits for student learning and teacher recruitment (Jung, Robinson, & Loeb, 2025).

Comprehensive monitoring will compare fidelity (dosage, instructional quality, retention) by tutor type, examining whether college students maintain comparable implementation quality. Implementation playbooks will document recruitment, training, and supervision requirements—enabling states to assess whether this workforce expansion approach offers viable scalability beyond traditional certified educator capacity.

National Significance and Field Contribution

Recent National Assessment of Educational Progress (NAEP) results demonstrate both capacity and need. Louisiana fourth graders achieved the largest reading gains nationally since 2019, rising from 50th to 16th in state rankings, demonstrating state capacity for evidence-based implementation at scale. Louisiana’s varying education landscape, 72 percent economically disadvantaged, nearly half rural, mirrors national challenges, making it the ideal testbed for statewide scaling. Findings will equip policymakers with evidence on three questions: (1) under what conditions virtual high-impact tutoring remains effective at scale; (2) whether extended dosage yields durable returns; and (3) how workforce models can sustain tutoring affordably.

B. Strategy to Scale

B.1. Unmet National and State Demand for Proven, Scalable Literacy Interventions

National Context: The Scaling Challenge

Across the U.S., ~78% of schools report offering literacy tutoring, yet only ~37% provide high-impact tutoring (HIT) that meets research standards for dosage, group size, and instructional quality (National Student Support Accelerator, 2025). As programs expand, dosage often declines and group size increases, eroding effectiveness. States need rigorous, policy-relevant evidence to answer:

- Which delivery models sustain effectiveness across rural, urban, suburban, and charter contexts?
- What implementation characteristics sustain quality at scale?
- How long should tutoring last to justify the cost (one vs. two years), and does the impact persist after tutoring ends?

This project directly addresses all three questions through a statewide, three-arm cluster RCT that directly tests dosage (one vs. two years), documents cost-effectiveness, and examines contextual moderators of impact.

Louisiana’s Documented Student Need

In Louisiana, 72.5% of students are economically disadvantaged, and 63 out of 64 parishes serve majority low-income populations. ~46% of K–6 schools are rural, well above national averages, creating access barriers that virtual delivery can uniquely address. The northeast Delta region faces significantly high poverty and will be prioritized.

Implementation Challenges from the first Statewide High-Impact Tutoring Initiative (Accelerate)

Statewide implementation through Louisiana's HIT initiative in 2023–24 school year produced positive outcomes, yet revealed three system-level gaps that this project is designed to resolve:

1. **Fidelity inconsistency across contexts.** Outcomes and adherence varied by provider, locale, and school type, with rural and charter schools facing distinct constraints. Controlled comparisons are required to identify models that sustain effectiveness across varied settings.
2. **Scheduling and operational barriers.** Protecting tutoring time in master schedules was difficult, especially in small schools. The state needs tested scheduling templates and evidence of which structures best maintain dosage.
3. **Cost sustainability concerns.** Districts need benchmarks on cost per student and cost per 0.10 SD gain, and evidence on whether a second year of tutoring yields durable gains that warrant the additional investment.

Readiness to Scale: Infrastructure and Evidence Foundation

Louisiana is uniquely prepared to scale, supported by the following existing initiatives:

- Universal screeners enable consistent progress monitoring (DIBELS).
- Science of Reading–aligned HQIM promotes instructional coherence.
- Professional learning networks reach educators statewide.
- The Accelerate: HDT policy framework mandates intervention and allocates funding.
- Effective HIT (Air Reading) that is backed by strong evidence meeting WWC standards.

These systems create the foundation for rigorous statewide testing and eventual scale-up.

Louisiana has the policy mandate to sustain tutoring beyond grant years, infrastructure to track student growth consistently, and the district readiness to participate in a large-scale trial. The proposed three-arm cluster RCT will leverage this existing capacity to generate policy-relevant

evidence on optimal duration, persistence of effects, and cost-effectiveness, producing a replicable statewide blueprint for scaling high-impact tutoring with fidelity and sustainability.

B.2. Management Plan

Louisiana will execute this multi-site implementation through a disciplined, milestone-driven management structure that leverages existing state infrastructure while ensuring fiscal accountability and continuous improvement. The plan combines clear governance, defined decision gates, experienced personnel, real-time data oversight, and proactive risk mitigation to deliver the project on schedule and within budget.

Governance and Accountability

- **Louisiana Department of Education's Division of Literacy and Tutoring** provides centralized leadership with defined roles and decision authority.
- **Project Director (LDOE)** manages daily operations, partner coordination, budget oversight, and federal compliance.
- **Monthly steering committee** (LDOE leadership, Project Director, JHU PI, Air Reading lead) reviews KPIs, resolves issues, and authorizes course corrections.
- **Quarterly SEA reviews** ensure alignment with the statewide literacy strategy and policy.

Timeline and Decision Gates

Phase One: Planning and Design (Calendar Year 2026)

- Finalize RCT design with JHU; preregister Pre-Analysis Plan (PAP); MOUs and DSAs; recruit 90 schools; site readiness; rural connectivity audits.
- **Decision Gate One (June 2026):** PAP pre-registered on OSF; MOU templates approved; 90 schools committed; data infrastructure specified.

Phase Two: Recruitment and Onboarding (July 2026–June 2027)

- MOUs executed; Air Reading contracts; data pipelines and dashboards; coordinator and tutor training; low-bandwidth configs; baseline data collection.
- **Decision Gate Two (June 2027):** 90 signed MOUs; tutor pool staffed (+15% reserve); coordinators trained; baseline complete; randomization verified.

Phase Three: Implementation Year One (School Year 2027–28)

- Launch in one-year and two-year arms; weekly fidelity monitoring; mid-year checks; targeted TA; corrective action for sites <80% fidelity.
- **Decision Gate Three (June 2028):** ≥80% schools meet dosage fidelity; ≤15% attrition; interim impact brief completed.

Phase Four: Implementation Year Two (School Year 2028–29)

- Continue two-year arm; one-year arm returns to business-as-usual; collect Y2 outcomes; document cross-context variation; initiate cost analysis.
- **Decision Gate Four (June 2029):** Y2 data collected; implementation documentation complete; cost-effectiveness in progress; draft playbooks.

Phase Five: Follow-up and Final Reporting (School Year 2029–30 and Fall 2030)

- One- and two-year follow-up; persistence analyses; finalize cost-effectiveness; WWC-standard reporting; toolkits; National Tutoring Policy Convening.
- **Decision Gate Five (June 2030):** One-year follow-up complete; preliminary persistence findings; playbooks finalized.
- **Decision Gate Six (December 2030):** Two-year follow-up complete; comprehensive final report; public deliverables released; replication support launched.

Key Personnel and Qualifications

Louisiana Department of Education (Lead Agency)

- [REDACTED] (0.1 FTE) – [REDACTED], LDOE.

Strategic leadership; alignment with statewide literacy framework; >experience managing federal literacy grants.

- [REDACTED] (0.3 FTE) – [REDACTED], LDOE. Leads operations, partner coordination, budget, timelines, reporting, and compliance.

Air Reading (Implementation Partner)

- [REDACTED] (0.2 FTE) – [REDACTED], Air Reading. Oversees statewide implementation strategy, tutor recruitment, fidelity monitoring, and platform scalability.

- Air Reading commits 5.5 FTE dedicated project staff, including a full-time [REDACTED]
[REDACTED]
[REDACTED] team members supporting operations across 90 schools.

Johns Hopkins University - CRRE (Independent Evaluator)

- [REDACTED] (0.16 FTE) – Principal Investigator, JHU-CRRE. [Redacted]
[REDACTED], leads RCT design, PAP, fidelity and impact analyses, cost-effectiveness, and reporting; [REDACTED].
- [REDACTED] (0.15 FTE) - Co-Investigator, JHU-CRRE. [REDACTED], specializes in educational psychology, teacher development, special education, mixed methods research of K-8th instructional interventions, and leads the [REDACTED]
[REDACTED] and is an [REDACTED]
- [REDACTED] (0.05 FTE) - Co-Investigator, CRRE. [REDACTED], textbook author, [REDACTED], and

technical advisor and researcher on federal and state initiatives involving evaluations of principals, technology, and learning outside of school.

Studyville Academic Workspace (Curricular/Community Partner)

- [REDACTED] (0.2 FTE) – [REDACTED]. Leads curriculum alignment to Louisiana's Guidebooks ELA, develops knowledge-rich writing components, coordinates instructional coherence with LDOE, and cultivates local partnerships.
- Studyville staff commits 1.2 FTE to support community engagement, curriculum development, and context-specific implementation guidance for rural and charter schools.

Risk Mitigation

Table 1 outlines the project’s plan for mitigating risk.

Table 1. Risk Mitigation Strategies		
Risk	Potential Impact	Mitigation Strategy
Tutor capacity shortfall	Reduced dosage; fidelity risk	15% reserve pool; rolling cohorts; retention incentives
Scheduling conflicts	Below-target dosage	State scheduling playbook (three tested templates); TA; protected time codified in MOUs
Rural connectivity issues	Missed sessions	Pre-launch audits; low-bandwidth configuration; asynchronous backups; device lending
Site staff turnover	Disruption; lower fidelity	Rolling onboarding; video training library; backup coordinators; quarterly re-engagement
Data integration delays	Monitoring gaps	Automated ETL to state dashboards; weekly data checks; site IT liaisons; escalation protocol

Monitoring and Continuous Improvement

- **Weekly:** Analytics (attendance, dosage); alerts for students/schools below thresholds.

- **Monthly:** Site observations for instructional quality; corrective action plans within 30 days for sites <80% fidelity.
- **Quarterly:** Statewide aggregation of trends (rural vs. urban; charter vs. traditional); steering committee reviews KPIs and authorizes adaptive supports; interim briefs to districts. Financial reports tracking expenditures vs. budget; variance analysis.
- **Annually:** Comprehensive progress reports (implementation, outcomes, costs); statewide convenings for feedback.

Key Performance Indicators (KPIs)

- **Access & enrollment:** $\geq 90\%$ of eligible students are scheduled by week three after the beginning of the year assessment.
- **Dosage:** $\geq 80\%$ of students receive prescribed dosage. $\geq 80\%$ student attendance rate.
- **Instructional fidelity:** $\geq 80\%$ of schools meet quality thresholds by mid-year (January); $\geq 85\%$ by year-end (May).
- **Recruitment/retention:** 90 MOUs by June 2027; ~ 50 students/school at baseline; $\leq 15\%$ attrition during implementation.
- **Impact:** Statistically significant gains on DIBELS/LEAP ELA; subgroup effects for high-need students.
- **Cost-effectiveness:** Annual estimates of cost per student and cost per 0.10 SD literacy gain; benchmarks published.

B.3. Partner Relevance and Commitment

Louisiana Department of Education (Lead Applicant). Provides statewide governance, fiscal oversight, policy integration, and data infrastructure. LDOE has demonstrated capacity scaling literacy initiatives statewide: Louisiana fourth graders achieved the nation's largest NAEP

reading gains since 2019 (rising from 50th to 16th), and K–3 students showed 17-point growth on universal literacy screeners in 2024–25. Commits 10% state match, priority access to literacy networks (e.g., Teaching & Learning calls to 1,200+ educators), and integration with Teacher Leader Summit. LDOE dedicates staff time and leverages proven capacity implementing federal literacy grants across all 64 parishes.

Local Education Agencies and Public Charter Schools (Sites). Competitively selected for readiness and commitment, 90 schools spanning urban, suburban, town, and rural contexts execute site-level implementation: scheduling, family engagement, logistics, monitoring participation, and data provision; site coordinator designation; protected intervention time in master schedules; data-sharing; and family engagement supports.

Air Reading (High-Impact Tutoring Implementation Partner). Delivers structured literacy tutoring model backed by strong evidence meeting WWC standards. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] Air Reading is responsible for tutor recruitment/training/supervision, and real-time fidelity monitoring. The virtual platform adheres to the Science of Reading, supports low-bandwidth environments, and delivers small-group instruction at prescribed dosage.

Johns Hopkins University-CRRE (Independent Evaluator). With established expertise in early literacy research, high-impact tutoring evaluation, and statewide implementation studies—including prior Air Reading evaluation in Louisiana—JHU-CRRE designs and executes the three-arm cluster RCT; conducts fidelity, moderator, and cost-effectiveness analyses;

measures persistence via one- and two-year follow-ups; and produces WWC-standard reports, technical appendices, and public-use datasets.

Studyville Academic Workspace (Curricular / Community Partner). With 20 years of experience in Louisiana education and deep knowledge of the state's curriculum landscape, Studyville develops knowledge-rich writing curriculum components aligned to Louisiana's HQIM (Guidebooks ELA), partners with LDOE to ensure instructional alignment and policy coherence, cultivates local community partnerships statewide, and advises on implementation for rural and charter contexts. Studyville will provide supplementary family engagement or community partnerships without modifying Air Reading's core tutoring curriculum.

These partnerships create an integrated implementation ecosystem: state policy leadership (LDOE), proven instructional delivery (Air Reading), independent validation (JHU-CRRE), community engagement (Studyville), and workforce expansion through Louisiana university partnerships provide the essential capacity for scaling. All partners have committed dedicated resources and staff time to ensure coordinated implementation and build sustainable capacity beyond the grant period.

B.4: Plan to Deliver Services Efficiently at Scale While Maintaining Effectiveness

Efficiency Through Strategic Design

Louisiana's delivery model achieves efficiency at scale while maintaining high-fidelity, evidence-based literacy instruction through high-impact virtual delivery. [REDACTED]

[REDACTED]

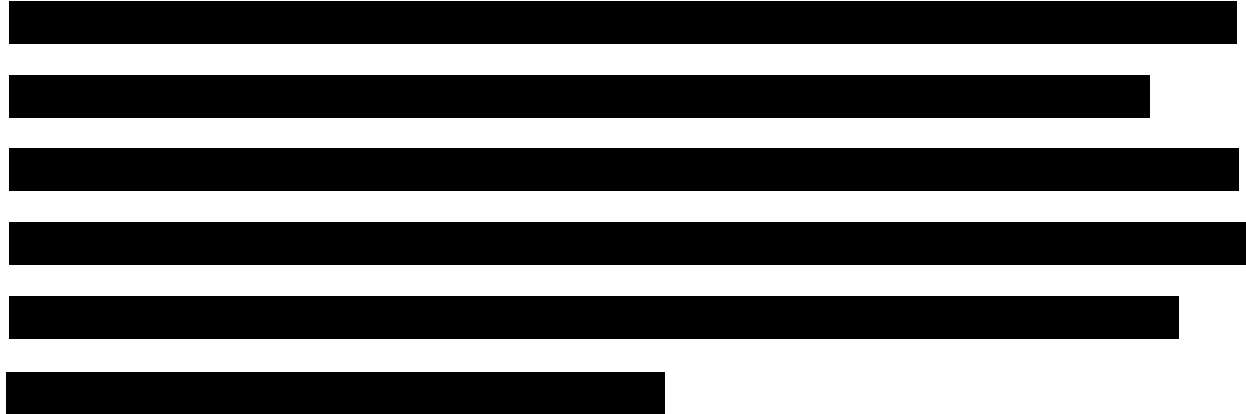
[REDACTED]

[REDACTED] The Air Reading program design balances standardization (ensuring fidelity) with local flexibility and student personalization (enabling

adaptation), as outlined in Table 2.

Table 2. Standardization and Flexibility		
Element	Standardized (ensures fidelity)	Flexible (enables local adaptation)
Instructional content	Purpose-built virtual tutoring structured literacy curriculum; research-based (Science of Reading, Orton-Gillingham); explicit, systematic instruction	Adaptive pacing based on student progress; supplementary materials aligned to local context
Progress Monitoring	DIBELS (BOY/MOY/EOY); Diagnostic; daily platform analytics; ongoing progress monitoring	School data review meeting; weekly student progress report; personalized instructional path
Grouping	1:2–1:3 ratios; consistent tutor assignment	Grade/skill-level grouping; mid-year regrouping based on progress
Dosage	Four sessions per week × 30 minutes × 20–24 weeks (≥60 sessions annually)	Scheduling format (block, ELA-adjacent, flexible); start windows; optional booster sessions
Platform	Air Reading digital platform; centralized monitoring; low-bandwidth optimization	Device type (laptops, tablets); location (classroom, lab, library); backup protocols

Systematic Quality Assurance



Maintaining Effectiveness at Scale

Multi-frequency assessment enables rapid instructional adjustments: daily mastery checks inform next-session instruction; weekly reports shared with classroom teachers enable integration of tutoring progress into core instruction; bi-weekly and monthly assessments track progress between DIBELS administrations. Multi-level monitoring (weekly analytics, monthly observations, quarterly statewide reviews) triggers targeted technical assistance and 30-day corrective action plans for low-fidelity sites.

This infrastructure, combining Air Reading's proven systems with state coordination, ensures consistent, high-fidelity implementation across Louisiana's varying contexts.

B.5. Dissemination Mechanisms

Louisiana will translate project findings into open, actionable resources enabling rapid replication by other states and districts. Dissemination emphasizes accessibility, scalability, and evidence translation, ensuring that both technical and non-technical audiences can adopt proven practices.

Tiered Product Strategy. The project produces a suite of publicly available deliverables tailored to replication needs:

- **SEA Blueprint for Statewide Tutoring Policy** – guidance for state leaders on funding models, accountability systems, and infrastructure requirements.
- **LEA Implementation Playbook** – turn-key adoption package including scheduling templates, fidelity rubrics, and family-engagement tools.
- **Technical and Research Outputs** – WWC-aligned evaluation reports, cost-effectiveness analyses, and public-use datasets on OSF and LDOE portals (detailed in Section D4).
- **Policy and Practice Briefs** – two- to four-page summaries translating evidence into funding, duration, and implementation recommendations for policymakers.

National Learning Collaborative. Beginning in 2026, Louisiana and Johns Hopkins will lead a multi-state Tutoring Learning Collaborative with 8-10 SEAs. Quarterly virtual sessions and annual summits share implementation tools, real-time data, and problem-solving strategies. By Year 4, 3-5 states receive intensive technical assistance, including MOU templates, cost modeling, and implementation coaching, preparing them for immediate adoption upon project completion.

Evidence-to-Policy Translation. Findings will be translated through three channels optimizing reach and usability: (1) Executive summaries distill technical findings into dosage recommendations and cost benchmarks for legislative audiences; (2) Practitioner briefs provide actionable implementation guidance for district leaders; (3) Conference presentations and webinars (AERA, CCSSO, SXSW EDU, National Title I Conference) enable direct engagement with implementation teams. These products help states determine optimal tutoring duration and budgeting models based on verified effectiveness data.

Open Access and Long-Term Availability. All resources, including datasets, codebooks, playbooks, and policy briefs, will be posted on the LDOE portal and Open Science Framework

(OSF) for permanent public access. Materials will carry Creative Commons licensing (CC BY 4.0) to encourage adaptation and reuse.

National Tutoring Policy Convening and Replication Launch (2030). A National Tutoring Policy Convening in Baton Rouge will synthesize five years of evidence and release the Replication Toolkit—final MOUs, cost models, vendor evaluation rubrics, and implementation guides. This event marks the formal launch of replication support for participating states and districts nationwide.

Together, these mechanisms ensure that Louisiana’s findings not only advance the research base but also establish a practical, replicable model for state-led, high-impact tutoring embedded within coherent literacy systems nationwide.

C. Quality of Project Design

C.1. Logic Model Framework

Louisiana's EIR Expansion project scales a proven, evidence-based virtual tutoring model to determine how Strong Evidence interventions perform under authentic statewide conditions. Foundational literacy acceleration for below-benchmark students requires intensive, structured-literacy tutoring delivered with high fidelity across varying contexts. Prior RCTs in Texas and Louisiana demonstrated statistically significant literacy gains, with stronger effects under full-year implementation ([REDACTED]). This expansion will implement the model in 90 schools while experimentally testing dosage (one vs. two years) to identify optimal and sustainable investment strategies. The logic model (Appendix G) links inputs → activities → outputs → outcomes to produce evidence that is both rigorous and policy-relevant.

Inputs and Their Causal Role

The project integrates five reinforcing inputs that establish conditions for scalability and sustained impact:

State Literacy Infrastructure. Louisiana's Science-of-Reading framework—including HQIM, universal DIBELS screening, professional learning networks, and dedicated tutoring funding—creates coherence between tutoring and Tier 1 instruction and supports continuity beyond grant funding.

Representative School Sample. Ninety elementary schools (urban, suburban, town, rural, charter) provide deliberate contextual variation, enabling moderator analyses to identify boundary conditions and contexts predicting the strongest implementation.

Evidence-Based Tutoring Platform. Air Reading delivers explicit, systematic instruction in phonemic awareness, phonics, fluency, vocabulary, and comprehension in 1:2–1:3 groups. Virtual delivery removes geographic barriers (~46% rural K-6 schools) and enables automated fidelity monitoring and real-time instructional adjustment.

Independent Evaluation Capacity. Johns Hopkins University–CRRE leads a five-year cluster RCT (MDES ~0.11–0.13) using DIBELS (short-term) and LEAP ELA (long-term) to measure impact, moderators, fidelity–outcome links, and cost-effectiveness (cost per 0.10 SD gain).

Workforce Expansion Pipeline. College students from Louisiana universities serve as tutors alongside professionals through work-study and service-learning partnerships—creating an abundant, affordable workforce and potential teacher pipeline.

Project Phases, Activities, and Outputs

Phase 1 (2026): Planning & Design. Finalize RCT design; pre-register PAP on OSF; recruit 90 schools; execute MOUs/DSAs; complete rural connectivity audits → PAP posted, data systems integrated, and WWC preconditions established.

Phase 2 (2026–27): Recruitment & Training. Prepare tutors through Air Reading's Teacher Academy (Science of Reading + platform training); finalize master schedules with protected tutoring blocks; collect baseline DIBELS → trained workforce and baseline data for ITT models.

Phase 3 (2027–28): Implementation Year 1. Deliver 4×/week, 30-minute tutoring for 20–24 weeks in one-year and two-year arms; monitor fidelity via platform analytics and observations; collect mid-year and end-year DIBELS → high-dosage delivery, fidelity attainment, and short-term impacts by workforce model.

Phase 4 (2028–29): Implementation Year 2. Continue two-year arm; collect cumulative outcome and cost data; document context variation → evidence on extended dosage, context moderators, and cost-effectiveness benchmarks.

Phase 5 (2029–30): Follow-Up & Dissemination. Conduct one- and two-year post-follow-ups; finalize analyses; publish implementation playbooks, cost benchmarks, and open datasets → persistence evidence, replication tools, and WWC-aligned public reporting.

Causal Pathways: How Inputs Produce Outcomes

Four testable causal pathways link project activities to measurable outcomes:

Cross-Context Implementation → Generalizable Evidence. Implementing across 90 schools tests replication across authentic state conditions and identifies contextual moderators of effect size. Outcome: external validity and context-specific implementation guidance.

Three-Arm RCT Design → Optimal Dosage Guidance. Random assignment to control, one-year, or two-year tutoring isolates causal effects of extended dosage and measures persistence after tutoring ends. Outcome: evidence determining whether two years justifies added investment based on cost-per-sustained-gain analysis.

System Integration → Sustainability Evidence. Embedding tutoring within Louisiana's Science of Reading framework (HQIM, DIBELS, PL networks) creates instructional coherence and skill transfer; integration within existing systems enables continuation beyond grant funding. Outcome: evidence on sustainable implementation and district continuation plans supported by state policy embedding HIT in MTSS.

Workforce Expansion → Scalability Evidence. College-student tutors supplement professionals through work-study partnerships, testing whether a lower-cost labor force maintains instructional quality and produces cost-efficient replication. Outcome: benchmarks on workforce composition, cost, and fidelity; validated university-district partnership model.

Outcomes Progression

Short-Term (2027–2029): Students receiving tutoring demonstrate statistically significant improvements in foundational literacy (DIBELS composite; target ≥ 0.11 SD vs. control). At least 80% of schools meet fidelity thresholds; effects disaggregated by dosage and context.

Mid-Term (2029–2030): Cost-effectiveness benchmarks established (cost per student and per 0.10 SD gain by dosage and workforce). Analyses identify practices supporting the strongest implementation across locales and confirm optimal dosage configuration. Implementation playbooks and toolkits published for statewide and national replication.

Long-Term (2030 and Beyond): One- and two-year follow-ups document persistence of literacy gains ($\geq 50\%$ of initial effects sustained). Louisiana embeds high-dosage tutoring within MTSS

through state policy and dedicated funding, continuing the Accelerate HDT initiative. At least half of participating districts sustain programming using existing budgets informed by cost benchmarks. At least 3–5 states initiate replication with technical assistance; peer-reviewed publications and open datasets enable national evidence synthesis.

Alignment with EIR Expansion Purposes

This logic model directly advances EIR Expansion's purpose: scaling Strong Evidence interventions while generating actionable evidence on dosage, persistence, cost-effectiveness, and sustainability. By connecting rigorous randomized evidence to practical implementation guidance, the project creates a state-led model that directly informs federal and state investment decisions, enabling nationwide replication of effective literacy tutoring at scale.

C.2. Goals, Objectives, and Outcomes

The project pursues four goals aligned with EIR Expansion's emphasis on scaling proven interventions while generating evidence on dosage, persistence, cost-effectiveness, and sustainable implementation. Each goal includes quantified performance measures, explicit objectives, and a rationale for being ambitious yet achievable.

Goal 1: Demonstrate Effectiveness at Scale

Primary Outcome: Statistically significant literacy gains (≥ 0.11 SD on DIBELS composite vs. control) by June 2028; $\geq 50\%$ sustained at one-year follow-up.

Why Ambitious Yet Achievable: Prior RCTs showed $+0.12$ SD overall and $+0.20$ SD in Grade 1 (██████████) and $+0.31$ SD full-year in Louisiana (██████████); powered sample (90 schools, ~4,500 students) and statewide infrastructure support fidelity.

Objectives:

- **1.1** Deliver tutoring to ~4,500 students (Grades 1–2, below proficiency) across 90 schools with $\geq 80\%$ completing ≥ 60 sessions annually (2027–29).
- **1.2** Achieve ITT effects meeting/exceeding MDES ~ 0.11 – 0.13 using DIBELS and LEAP ELA with 95% CIs.
- **1.3** Publish annual fidelity dashboards on LDOE and OSF portals (2027–30).

Goal 2: Measure Persistence and Determine Optimal Dosage

Primary Outcome: $\geq 50\%$ of immediate effects sustained at one- and two-year follow-ups; cost per 0.10 SD gain by dosage configuration.

Why Ambitious Yet Achievable: Early foundational skills and alignment to core instruction support durability; JHU cost expertise enables rigorous ingredients-based analysis.

Objectives:

- **2.1** Compare one- vs. two-year arms through pre-registered cluster RCT measuring persistence via DIBELS and LEAP follow-ups (2027–30).
- **2.2** Calculate cost per student and cost per 0.10 SD gain by dosage and workforce model (by December 2029).
- **2.3** Identify optimal dosage strategy and publish policy briefs translating findings for state budget decisions (2030).

Goal 3: Achieve Implementation Fidelity at Scale

Primary Outcome: $\geq 80\%$ dosage fidelity; $\geq 70\%$ adherence to core components; average quality $\geq 3/4$ on standardized rubrics (SY 2027–29).

Why Ambitious Yet Achievable: Challenging across varied contexts; achievable through coaching, technical assistance, and virtual guardrails; Louisiana operational experience (102 schools, 2024–25) validates feasibility.

Objectives:

- **3.1** Monitor dosage, adherence, and quality via platform logs and monthly observations; implement 30-day corrective action plans for low-fidelity sites (2027–29).
- **3.2** Establish state infrastructure (MOUs, data pipelines, protected intervention blocks) for 90 schools by June 2027.

Goal 4: Determine Cost-Effectiveness and Build Replication Capacity

Primary Outcome: Published cost benchmarks; $\geq 50\%$ of districts draft continuation plans; 3–5 states initiate replication; replication toolkits released (by December 2030).

Why Ambitious Yet Achievable: LDOE's Accelerate HDT infrastructure and dedicated funding demonstrate sustained capacity; practical toolkits ease adoption; workforce expansion through college students provides scalable capacity.

Objectives:

- **4.1** Recruit and train ~200 tutors (educators + college students) annually maintaining $\geq 85\%$ retention and $\geq 80\%$ fidelity (2027–29).
- **4.2** Document cost differentials and quality outcomes by tutor type; produce replication toolkits, cost benchmarks, and open datasets (2029–30).
- **4.3** Support 3–5 partner states through National Learning Collaborative with intensive technical assistance (2028–30).

Measurement and Analysis

All goals are measured through JHU-CRRE's independent evaluation meeting WWC standards without Reservations. Three-arm cluster RCT with pre-registered analysis plan, intent-to-treat models, and 95% confidence intervals ensures rigor. Continuation plans will specify funding sources (Title I, state literacy funding, local budgets), ensuring tutoring becomes a permanent component of Louisiana's literacy strategy beyond federal grant support.

D. Project Evaluation or Other Evidence-Building

D.1. Methods of Evaluation Will Produce WWC-Standard Evidence

Overview. An independent cluster randomized controlled trial (JHU-CRRE) will test whether **one year vs. two years** of Air Reading tutoring yields **significant, durable** literacy gains relative to business-as-usual (BAU). The design, measures, analysis, and reporting are configured to meet **WWC standards without Reservations** and to answer policy-relevant questions on **effectiveness, optimal dosage, persistence, and cost-effectiveness**.

Study Design and Random Assignment

- **Design:** Three-arm, school-level cluster RCT (to prevent contamination and mirror real staffing/scheduling).
 - **Arm A (BAU):** Continues standard literacy instruction/supports.
 - **Arm B (1-Year Tutoring):** Tutoring in 2027–28; then BAU while persistence is measured.
 - **Arm C (2-Year Tutoring):** Tutoring in 2027–28 and 2028–29 for cumulative effects.
- **Sample:** 90 schools (~50 students each; total sample ~4,500 students) spanning rural, urban, suburban, traditional, and charter schools.

- **Timing:** Planning/instrument refinement (2025–26); recruitment/pilot (2026–27); implementation (2027–29); follow-up (2029–30).
- **Policy comparisons enabled:**
 - **Year-1 impact:** Arm B & C vs. BAU (end of 2027–28).
 - **Year-2 cumulative impact:** Arm C vs. BAU (end of 2028–29).
 - **Dosage effect:** Arm C vs. Arm B (end of 2028–29).
 - **Persistence:** Both treatment arms vs. BAU at one- and two-year follow-ups.

Outcomes and Measures

To capture both foundational skill growth and transfer to grade-level proficiency, the evaluation employs two complementary outcomes summarized in *Table 3*.

Table 3. Primary Outcomes			
Outcome	What it captures	Windows	Rationale
DIBELS 8th Ed. composite	Foundational skills (phonemic awareness, decoding, fluency) targeted by Air Reading	BOY/MOY/EOY in 2027–29; 1-yr follow-up (Spring 2030)	Sensitive to short-term skill growth; widely used; strong psychometrics; aligns with LA early literacy standards
LEAP 2025 ELA	Reading comprehension and proficiency	Spring window(s) through 2029 and at follow-ups	Policy-relevant, authentic transfer to grade-level text;

Both outcome measures satisfy WWC requirements, with established validity and reliability, uniform administration across arms, and aligned, pre-specified measurement windows.

Sample and Statistical Power

To ensure the study is adequately powered to detect policy-relevant effects, JHU used PowerUp! (Dong et al., 2013) with ICC and R² assumptions drawn from Hedges & Hedberg (2007). The

design specifies $\alpha = .05$ (two-tailed) and power = .80, with an average cluster (school) size = 50 students and an assumed 15% student-level attrition (no school attrition). Under these parameters, the minimum detectable effect sizes (MDES) for intent-to-treat estimates are +0.11 SD for the Year-1 comparison (Arm B vs. BAU), +0.13 SD for the Year-2 comparison (Arm C vs. BAU), and +0.14 SD for persistence analyses. Because these MDES values are smaller than the +0.12 SD overall effects and +0.20 SD Grade 1 effects observed in prior Air Reading RCT work, the study is well-positioned to detect impacts if they replicate under statewide conditions.

Analysis Plan

All analyses will be pre-registered on OSF prior to baseline. The primary framework is an intent-to-treat hierarchical linear model with students nested in schools, school random effects, and baseline covariates (prior literacy scores, grade, demographics) for precision. JHU will estimate three pre-specified contrasts: (1) Year-1 impact B & C vs. BAU, (2) Year-2 cumulative impact C vs. BAU and incremental C vs. B, and (3) persistence at one- and two-year follow-ups. Results will report effect sizes, 95% confidence intervals, and p-values. Pre-specified moderator analyses will test heterogeneity by baseline proficiency (below vs. well-below), SES, EL status, race/ethnicity, grade, urbanicity, charter vs. traditional status, baseline resources, and tutor type. Finally, exploratory fidelity–outcome linkages will examine whether higher dosage, adherence, or instructional quality predict larger gains, identifying the active ingredients most critical for replication.

Threats to Validity and Mitigation

School-level random assignment eliminates selection bias, ensuring treatment and control groups are equivalent in expectation on all measured and unmeasured characteristics. Randomization at

the school level prevents within-school contamination where treatment students might influence control students or teachers might adjust instruction based on student assignment.

Student and school attrition will be minimized through strong district partnerships, maintaining consistent communication with participating schools, proactive tracking of student mobility, and data monitoring throughout implementation. If differential attrition emerges (higher rates in one condition), sensitivity analyses including bounding procedures will assess the robustness of findings. Baseline equivalence will be verified statistically; if imbalances emerge despite randomization, sensitivity analyses will test whether results are robust to covariate adjustment.

Implementation variation will be monitored through session-level fidelity data and observation protocols. These data enable interpretation of null findings (was intervention ineffective, or was it not delivered as designed?) and link implementation quality to outcomes.

The evaluation design maximizes both internal validity (random assignment, ITT analysis, pre-registration) and external validity (varied real-world implementation contexts, policy-relevant research questions, authentic outcome measures), producing credible causal evidence with clear applicability to state literacy policy decisions.

D.2. Evaluation Will Provide Guidance for Replication and Testing in Other Settings

The evaluation generates actionable guidance enabling state and district leaders to replicate the model successfully by testing implementation across varied contexts, documenting operational requirements in detail, and examining conditions predicting success. Rather than asking only "does it work?", the study answers practitioners' critical questions:

- What resources are required?

- Which contexts enable the strongest implementation? What conditions are necessary for effectiveness?
- What costs should we expect?

Testing Across Varied Implementation Contexts

Louisiana's 90-school sample provides natural variation across critical contextual dimensions. Schools span rural communities (where geography creates access barriers and workforce challenges), urban districts (with large enrollments and complex scheduling demands), and suburban settings (typically with stronger resource bases). The sample includes traditional public schools operating under district governance and public charter schools with independent operational flexibility. Schools vary in per-pupil spending, staffing levels, baseline literacy proficiency, student demographics, and existing intervention infrastructure. Baseline analyses will document how participating schools compare to Louisiana schools statewide on demographics, urbanicity, and resources to clarify generalization boundaries.

This variation enables testing whether Air Reading's virtual model achieves consistent results despite varying local conditions—the central question for statewide scaling. Moderator analyses will examine whether impacts differ by urbanicity, charter status, baseline resources, and tutor type. Implementation documentation will capture what differs across contexts: Do rural schools face connectivity challenges? Do urban schools struggle with scheduling complexity? Do charters implement differently from traditional schools? Results identify facilitators and barriers specific to each context type, moving beyond average effects to context-specific guidance.

By simultaneously implementing across Louisiana's educational ecosystem rather than in carefully selected demonstration sites, the evaluation tests the model under authentic conditions where most schools will eventually implement, strengthening external validity and replicability.

Findings will show not just whether the model works on average, but where it works best, what adaptations enable success in challenging contexts, and which contexts may require additional supports for effective implementation.

D.3. Quality of the Fidelity of Implementation Plan

Implementation fidelity measurement is essential for both interpreting findings and enabling replication. If intervention produces effects, fidelity measurement confirms the model was delivered as designed. If effects are smaller than expected, fidelity data reveals whether implementation challenges explain results rather than intervention ineffectiveness. For replication, fidelity measurement identifies which components are essential and what implementation standards must be maintained.

Core Components and Active Ingredients

The Air Reading tutoring model comprises four essential components hypothesized to produce literacy gains. Together, these constitute the active ingredients hypothesized to accelerate foundational literacy skills.

Dosage and structure: 4 sessions weekly, 30 minutes per session, 20–24 weeks annually, in groups of 1:2–1:3, providing intensity and individualization impossible in typical classrooms.

Instructional content: Structured literacy curriculum aligned to Science of Reading with systematic scope and sequence targeting phonemic awareness, phonics, fluency, vocabulary, and comprehension using explicit, systematic instruction.

Tutor quality: Trained tutors receiving systematic coaching with structured observation and actionable feedback, maintaining instructional quality.

Progress monitoring: Ongoing assessment enabling instructional adjustments responsive to student needs.

Fidelity Measures and Pre-Specified Thresholds

Dosage fidelity will be measured through automated platform logs documenting sessions delivered per student, with a threshold of at least 80% of scheduled sessions completed (minimum 65 sessions over 20–24 weeks). This threshold reflects prior research showing dose-response relationships where insufficient dosage is unlikely to produce effects. Platform data captured continuously enables weekly monitoring with monthly site-level summaries triggering corrective action when schools fall below thresholds.



Using Fidelity Data for Interpretation and Improvement

Fidelity data serves dual purposes, enabling both rigorous interpretation and continuous improvement. For interpretation, comprehensive fidelity documentation allows WWC reviewers and practitioners to distinguish "intervention doesn't work" from "intervention wasn't implemented as designed." If effects are smaller than expected, fidelity analyses test whether implementation quality explains results.

For improvement, real-time monitoring enables rapid correction. Monthly fidelity reviews identify schools below thresholds; corrective action plans provide targeted support, including

additional tutor coaching, schedule adjustments protecting intervention time, or technical assistance addressing connectivity issues, preventing sustained low-quality implementation that would undermine effectiveness.

For analysis, fidelity measures will be incorporated as moderators in exploratory models testing whether stronger implementation predicts larger effects, clarifying which aspects of implementation—dosage adherence, instructional quality, or specific components—most critically drive outcomes and must be prioritized in replication.

D.4. Design Will Generate Information to Guide Replication and Scaling

The evaluation produces evidence on cost-effectiveness, persistence, and implementation conditions essential for sustainable scaling decisions. Beyond measuring impacts, the study answers questions enabling states and districts to determine whether and how to adopt the model.

Comprehensive Cost and Cost-Effectiveness Analysis

Ingredients-based cost analysis will document all resources required: personnel costs (tutor salaries/stipends, training time, coaching staff); technology costs (platform fees, devices, connectivity infrastructure); materials costs (curriculum, assessment tools); administrative costs (coordination, data management, oversight); and opportunity costs (student time, instructional space). Both direct costs paid from grant funds and indirect costs, including in-kind contributions, will be captured. Costs will be tracked separately for one-year versus two-year arms, enabling direct dosage comparison.

Cost-effectiveness metrics will include cost per student served, cost per 0.10 standard deviation literacy gain, and literacy gains per \$1,000 invested, enabling comparison to other literacy interventions. Break-even analysis will identify the effect size at which an extended two-year dosage justifies additional investment compared to a one-year intervention, directly

informing resource allocation decisions. These analyses provide actionable benchmarks for sustainable budgeting using state and local funds after federal grant support ends.

Persistence: Evidence on Durability

Follow-up measurement one and two years after tutoring ends provides critical evidence on durability—whether literacy gains are sustained when intensive support is removed. Persistence analyses compare treatment arms to control at follow-up timepoints, calculating persistence ratios (follow-up effect size divided by immediate post-intervention effect size), showing what proportion of initial gains remain. If one-year tutoring produces equally persistent gains as two-year tutoring at a lower cost, it may be the optimal investment. If two-year tutoring shows superior persistence justifying additional expense, extended investment may be warranted. Results inform not just whether intervention works, but optimal duration for a durable impact.

Supporting National Replication

Findings will be disseminated through comprehensive WWC-aligned technical reports, peer-reviewed publications, practitioner briefs translating results for non-technical audiences, and an open-access data repository with codebooks and analysis code on Open Science Framework, enabling independent verification. Implementation playbooks with scheduling templates, fidelity monitoring tools, and cost benchmarks will provide turn-key guidance for adopting states and districts. Public transparency strengthens evidence credibility and accelerates knowledge accumulation nationwide.