

**Arkansas Statewide Rural Literacy Study: Testing Evidence-Based High-Impact Tutoring at Scale**

*Table of Contents*

<b>A. Significance</b>	<b>1</b>
A.1. The Rural Literacy Crisis and Proven Solutions	2
A.2. Four Research Innovations - Replication. Scale. Learning. Workforce.	3
A.3. National Significance and Field Contribution	5
<b>B. Strategy to Scale</b>	<b>6</b>
B.1. Unmet National and State Demand for Evidence-Based Rural Literacy Solutions	6
B.2. Feasibility of the Management Plan	10
B.3. Relevance and Demonstrated Commitment of Each Partner	14
B.4. Plan to Deliver Project Services More Efficiently at Scale and Maintain Effectiveness	15
B.5. Quality of Mechanisms to Broadly Disseminate Information and Resources	17
<b>C. Quality of Project Design</b>	<b>18</b>
C.1. Logic Model: Quality of Conceptual Framework Linking Inputs to Outcomes	18
<b>D. Quality of Project Evaluation or Other Evidence-Building</b>	<b>25</b>
D.1. Methods of Evaluation and What Works Clearinghouse Standards	26
D.2. Evaluation to Provide Guidance for Replication or Implementation	29
D.3. Measuring Fidelity of Implementation and Linking to Outcomes	31
D.4. Design for Valid, Reliable Information to Guide Replication	33
<b>References</b>	

## Arkansas Statewide Rural Literacy Study: Testing Evidence-Based High-Impact Tutoring at Scale

### A. Significance

**Nearly 70% of rural fourth-grade students nationwide are performing below grade level**

(2024 NAEP results). Rural districts struggle to deliver evidence-based interventions at scale.

Geographic isolation, severe staffing constraints, and limited specialist capacity mean that even when funding exists, many rural communities cannot provide the targeted literacy support their students need. Efforts to close achievement gaps remain out of reach for many rural communities.

The proposed EIR Expansion project addresses this challenge by rigorously testing an evidence-based virtual tutoring model (Air Reading) across 100 diverse Arkansas rural schools. Through a large-scale Randomized Controlled Trial (RCT), Arkansas will answer four questions every rural state urgently faces:

1. **Replication:** Do evidence-based programs from other states work in our context—with different standards, assessments, and curriculum?
2. **Scale:** How do we effectively scale a proven model across 100 diverse rural schools—and what changes are needed to make the model work reliably?
3. **Learning:** How do we invest early to ensure struggling first-graders reach proficiency by 3rd grade—the critical gateway to all learning?
4. **Workforce:** Given severe rural staffing shortages, can college students trained in Science of Reading deliver literacy gains comparable to experienced paraeducators—while exploring whether tutoring experience increases teaching interest?

By answering these questions through this RCT, Arkansas will produce actionable evidence and implementation resources—covering adaptation, scale, cost-effectiveness, and workforce strategies that other rural states can adopt immediately.

### **A.1. The Rural Literacy Crisis and Proven Solutions**

**Arkansas exemplifies the national rural literacy crisis.** Rural schools educate 54% of students statewide—60% of K–3 schools (Office for Education Policy, 2024, 2025). Among rural third-graders, 28% score Level 1 (far below grade level) on Arkansas Teaching, Learning & Assessment System (ATLAS), rising to 41% in Delta regions. Only about one-third reach proficiency. 64% of rural students qualify for free/reduced lunch; 3.7% experience homelessness (vs. 2.4% nationally; NCHE, 2023). Severe staffing shortages mean cooperatives often have one literacy specialist serving multiple campuses, and after-school models exclude homeless students who depend on bus transportation.

One evidence-based model has shown promise in addressing these rural constraints. The implementation partner, Air Reading, delivers a Science of Reading-based literacy intervention through virtual high-impact tutoring.

**Strong What Works Clearinghouse (WWC)-Standard Evidence Base:** An independent RCT of Air Reading in rural Texas (Neitzel & Storey, 2024) found statistically significant literacy gains on NWEA MAP Reading among Tier 2/3 students (72% economically disadvantaged; 13% special education; 25% bilingual), meeting WWC Group Design Standards without reservations. Students receiving  $\geq 40$  sessions per semester showed the largest effects. A follow-up RCT of Air Reading in suburban Louisiana (Neitzel & Storey, 2025;  $n=170$ ) found even stronger results ( $+0.31$  SD,  $p < .05$ ) with a full-year implementation, suggesting extended duration enhances outcomes.

This strong foundation positions Arkansas to extend and test the model statewide under real-world, state-led implementation conditions — answering whether proven effects transfer

when adapted to different standards, assessments, and infrastructure, and whether the model maintains quality when scaled across 100 diverse rural schools.

**A.2. Four Research Innovations - Replication. Scale. Learning. Workforce.**

Arkansas will implement four innovations that together answer whether evidence-based literacy tutoring can achieve reliable impacts when adapted and scaled across diverse rural contexts.

These innovations address critical implementation questions: cross-state transferability, quality at scale, optimal intervention duration, and workforce expansion.

**Replication - Adaptation Across Standards and Systems:** This study will adapt Air Reading to Arkansas’s context, distinct standards, assessments, and infrastructure. Texas’s RCT used NWEA MAP; Arkansas uses ATLAS Reading/ELA for accountability and DIBELS 8/FastBridge for early literacy monitoring. These differences require recalibrating placement criteria for intervention determination, progress thresholds, and exit decisions. Likewise, Arkansas ELA standards emphasize different sequences and grade-level expectations than Texas TEKS, and the state’s infrastructure varies dramatically, from multi-grade classrooms and early bus schedules to limited broadband capacity. Implementation will align with ADE's HIT Instructional Coherence Toolkit (July 2025) to strengthen Tier 1–3 integration.

Significance: External validity rarely transfers automatically. This design tests whether an evidence-based tutoring model proven elsewhere maintains its effects when adapted to Arkansas-specific assessments, standards, and infrastructure—offering crucial insight for other states with differing accountability systems and curricular frameworks.

**Scale - Statewide Implementation Across 100 Rural Schools:** This project expands from six Texas schools to 100 rural Arkansas schools, spanning all fifteen regional cooperatives and the state’s major rural regions (Delta, Ozarks, River Valley). Schools range from small campuses

(<100 students) to larger rural schools (300+ students) and from remote to rural-fringe locales. Some have specialists; others rely on teachers who cover multiple grades.

Significance: Few RCTs test both internal and external validity. This 100-school design does both: it captures the full range of real-world variation—staffing constraints, infrastructure challenges, geographic isolation. By testing across this diversity, the study identifies which implementation factors preserve effectiveness and which adaptations are necessary, generating the first large-scale evidence on implementation in rural schools, statewide.

**Learning - Pathways to Grade 3 Reading Proficiency:** Third-grade reading proficiency is the critical milestone—students below proficiency are four times more likely to drop out (Hernandez, 2011). States face a fundamental question: if we intervene with struggling first-graders, can we change their trajectory to reach proficiency by Grade 3?

Arkansas tests this by providing one year of intensive literacy intervention to Level 1 first-graders, then following up with a second year of intervention for a subset of students. By tracking outcomes through Grade 3 state assessments, the study examines whether early intervention produces lasting proficiency—the outcome that matters most for accountability.

Significance: States invest heavily in early literacy funding but lack evidence on whether early intervention produces Grade 3 proficiency—and whether one year or sustained two-year support is necessary. This study compares both durations through state assessments, reporting cost per proficient reader for each duration to enable data-driven investment decisions.

**Workforce - College Students as Literacy Tutors:** Severe rural staffing constraints limit tutoring reach even when funding exists. This embedded RCT tests whether trained college students can achieve literacy outcomes comparable to experienced paraeducators. Students in

treatment schools are randomly assigned to experienced paraeducators or college students, with both groups receiving identical Science of Reading training and coaching.

Significance: If college students perform comparably to experienced paraeducators, Arkansas can expand tutoring capacity rapidly and cost-effectively, leveraging work-study funding to address immediate rural workforce shortages. The study will also observe whether tutoring experience increases participants’ interest in teaching, building on emerging evidence that tutoring substantially increases teacher-preparation applications, particularly in non-education majors ( [REDACTED] )—offering insights into potential pathways for diversifying the teaching workforce.

### **A.3. National Significance and Field Contribution**

Rural education challenges are structurally similar across the country. States such as Montana, Wyoming, North Dakota, and South Dakota have 70–90% rural schools; Mississippi, West Virginia, New Mexico, and Idaho share similar poverty and staffing patterns; Alaska faces extreme geographic challenges. Arkansas represents a typical rural context, making it an ideal testbed for solutions that can scale nationally.

This study will produce two national resources: (1) **Rural State Literacy Tutoring Toolkit**, guiding adaptation, implementation, cost analysis, and policy alignment for evidence-based literacy tutoring in rural contexts; and (2) **College Student Tutoring Workforce Playbook**, covering recruitment, training, supervision, cost-effectiveness analysis, and findings on teaching interest among college student tutors. Together, these guides will enable rural states to implement a tested model for improving early literacy at scale, grounded in rigorous research and practical implementation evidence.

## **B. Strategy to Scale**

### **B.1. Unmet National and State Demand for Evidence-Based Rural Literacy Solutions**

With nearly 70% of rural fourth-graders reading below grade level nationally (NAEP, 2024), states invest billions annually in early literacy but lack rigorous evidence on implementation at scale in rural contexts. Rural states nationwide face common challenges—geographic isolation, severe staffing constraints, limited specialist capacity—yet must make high-stakes decisions about intervention adoption, duration, and workforce strategies without context-specific evidence. The proposed study addresses this national gap by testing replication, scale, duration, and workforce solutions in Arkansas—findings directly transferable to states with similar rural demographics and infrastructure.

Arkansas has demonstrated both urgent demand and strong commitment. With 217 eligible rural elementary schools serving over 3,200 first-graders scoring well below grade level (ADE 2025 projections), demand exceeds this 100-school study's capacity by 3–4 times. State leadership—Secretary Jacob Oliva, superintendents, regional cooperatives, and family advocates—have expressed strong interest in evidence-based solutions, indicating substantial, persistent, unmet demand from educators, policymakers, and communities.

**Evidence Gaps from Arkansas's \$33 Million High-Impact Tutoring Initiative:** Arkansas launched a statewide High-Impact Tutoring (HIT) initiative in January 2024, investing \$11 million that month, and another \$22 million in the 2024–25 school year. This launch demonstrates both demand and state commitment. However, implementation revealed critical evidence gaps. Arkansas invested \$33 million without rigorous data on cost-effectiveness, optimal duration, or workforce strategies—creating fiscal accountability challenges. Outcomes were varied, with some schools showing gains while others saw no impact; many never launched

successfully. Fidelity broke down at scale—sessions were canceled due to tutor absences; tutors taught content disconnected from Tier 1; after-school models excluded homeless students; and some tutors lacked Science of Reading training. Many districts redirected funds to existing staff rather than hiring dedicated tutors, diluting dosage and fidelity. Even with \$33 million available, rural districts could not recruit qualified tutors, and districts interested in college students lacked content or training protocols to do so safely and effectively.

These implementation challenges underscore the need for rigorous, context-specific evidence. Long-standing structural barriers persist regardless of funding: pull-out interventions often operate disconnected from Tier 1 instruction, specialists must serve multiple campuses, transportation constraints fragment after-school models, and few systems coherently align Tier 1–Tier 3 supports.

**How This 100-School Study Addresses Unmet Demand:** The proposed 100-school cluster RCT addresses unmet state and national demand through four targeted innovations responding to HIT's implementation challenges:

- Arkansas-specific adaptation: Aligns tutoring with state standards, ATLAS assessments, and the seven regional cooperative infrastructure—addressing HIT's challenge of disconnected content.
- Systematic implementation protocols: Sessions scheduled during the school day, integrates Tier 1–Tier 3 instructional coherence via ADE's HIT Toolkit, provides Science of Reading training, and uses virtual delivery to ensure access for all students, especially those experiencing homelessness—addressing HIT's fidelity and access challenges.

- Duration impact on learning: Tests one-year vs. two-year participation, measuring Grade 3 outcomes and cost-effectiveness—filling the fiscal accountability gap from \$33M HIT investment.
- Workforce solutions: Compares experienced paraprofessionals or certified teachers with college student tutors recruited through ADE–IHE partnerships—addressing HIT's recruitment challenges and testing scalable workforce expansion.

Regional education cooperatives, families, and local school leadership will serve as active implementation partners throughout the study, ensuring solutions reflect community needs and priorities. The project will produce two guides—*Rural State Literacy Tutoring Toolkit* and *College Student Tutoring Workforce Playbook*—enabling other states to implement the program while drawing upon the lessons learned from Arkansas's earlier initiative. Together, these strategies transform Arkansas's current investment into a structured, evidence-producing system that other states can replicate.

**Alignment Between Demand and Scale:** The 100-school scale directly addresses both immediate and broader implementation demand. With 217 eligible schools requesting participation—exceeding study capacity by 3–4 times—this design enables Arkansas to:

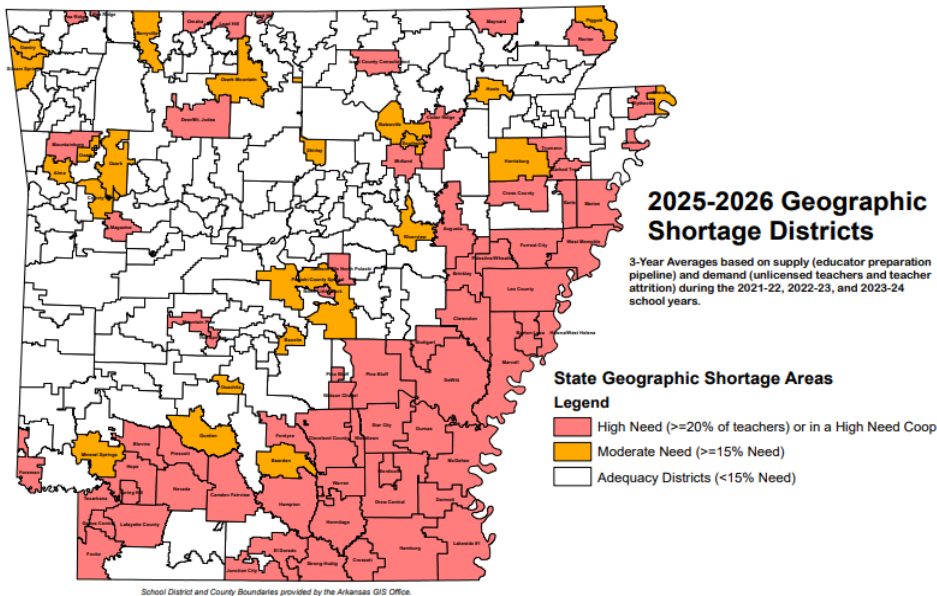
Generate rigorous evidence for statewide expansion: The 100-school pilot (50 treatment, 50 control) spans all seven regional cooperatives and Arkansas's major rural regions (Delta, Ozarks, River Valley), capturing the full diversity of school contexts (remote to rural-fringe; small to large; varying specialist capacity). This ensures findings are transferable to all 217+ eligible Arkansas schools seeking to implement after validation.

Produce actionable guidance for national adoption: By testing across diverse rural conditions, the study generates implementation protocols rural states nationwide can adopt—addressing persistent demand from states with similar rural demographics and infrastructure challenges.

Maintain quality and fidelity: The 100-school scale is large enough to demonstrate reliability across real-world variation yet focused enough to ensure rigorous implementation monitoring and WWC-compliant evidence—addressing the fidelity breakdowns that plagued Arkansas's earlier \$33M HIT initiative.

If oversubscribed, schools with the largest percentage of Level 1 students will be prioritized, directing resources to the most acute needs. Even after this study, unmet demand persists within Arkansas and across rural states—underscoring both the urgency and the potential for broader nationwide implementation once protocols are validated through this pilot.

**Figure 1. Arkansas 2025-26 Teacher Shortages by County and District**



Source: Division of Elementary & Secondary Education. “Geographic Shortage Areas.” *Educator Workforce*

*Data.* Arkansas Department of Education, Retrieved Oct. 8, 2025

## **B.2. Feasibility of the Management Plan**

ADE will manage this 100-school expansion through an existing statewide infrastructure already tested under its \$33 million High-Impact Tutoring (HIT) initiative launched in January 2024.

The management plan leverages ADE's proven statewide systems to ensure the project is implemented on time and within budget. Clear lines of authority, existing data pipelines, and regional cooperative infrastructure enable real-time oversight and rapid course correction.

### **Key Feasibility Metrics**

- $\geq 90\%$  of treatment schools launched within 2 weeks after the BOY assessment window
- $\geq 80\%$  students received prescribed dosage/student attendance rate
- $\geq 95\%$  of scheduled minutes delivered
- $\geq 90\%$  school retention year-to-year
- Quarterly KPI reviews and annual milestone audits conducted by ADE and Stanford

### **Governance, Accountability, and Monitoring**

Executive Oversight: [REDACTED], ADE Learning Services

- Project Director and EIR signatory. Responsible for compliance, budget alignment, and milestone delivery. Chairs quarterly Executive Steering Committee.
- Performance Standards: 100% milestones met on schedule; quarterly KPI reviews completed; budget variance  $< 5\%$  annually

Implementation Oversight: [REDACTED]

- Oversees day-to-day operations, school recruitment, and data reporting.
- Performance Standards:  $\geq 98\%$  data accuracy;  $\geq 90\%$  school satisfaction; all MOUs executed on schedule

Evaluation Oversight: [REDACTED], Stanford University

- Provides external validation, evaluation integrity, WWC alignment, and cost-effectiveness reporting.
- Performance Standards: 100% reports WWC-aligned;  $\geq 95\%$  data completeness; all deliverables submitted on time

Tutor Management: [REDACTED] Air Reading

- Oversees tutoring delivery, technical support, and real-time data dashboards. Reports weekly to ADE.
- Performance Standards:  $\geq 85\%$  median fidelity sustained;  $\geq 95\%$  platform uptime;  $< 3$  unresolved support tickets per month

IHE Partners Relations: [REDACTED], Air Reading

- Recruit, train, and supervise college tutors; field placement integration;
- Performance Standards:  $\geq 150$  college tutors recruited annually;  $\geq 50\%$  express teaching interest;  $\geq 80\%$  retention semester-to-semester

Regional Implementation: Fifteen Regional Cooperative Directors

- Provide regional coaching, fidelity monitoring, and school-level escalation. Each reports monthly KPI data to ADE.
- Performance Standards:  $\geq 85\%$  of schools meet fidelity thresholds;  $< 5\%$  session cancellations; 48-hour response time for school support

### **Monitoring, Reporting, and Continuous Improvement**

- Weekly Reviews: ADE Implementation Team reviews flags (e.g.,  $< 75\%$  dosage,  $< 80\%$  fidelity) and initiates outreach same week. Coach observations scored.
- Monthly Partner Coordination Meetings: ADE convenes monthly data reviews with Stanford to validate sample integrity and performance metrics

- Quarterly Executive Reviews: Summarize key performance indicators, risk status, budget utilization, and recommended course corrections
- Annual External Evaluation Reviews (Stanford): Stanford publishes annual reports summarizing interim outcomes, fidelity patterns, and operational refinements

<b>Table 1. Detailed Timeline with Milestones</b>		
<b>Phase &amp; Dates</b>	<b>Milestones &amp; Deliverables</b>	<b>Lead &amp; Supporting Roles</b>
Readiness & System Build (Jan–Jul 2026)	<b>Milestone 1 (July 2026):</b> 100 school MOUs executed (≥110 initially recruited to offset attrition). 5–8 IHE agreements signed. College recruitment plan approved by ADE. ≥ 20 pilot schools selected and cleared by Stanford for RCT eligibility.	<b>ADE</b> – MOU execution, data integration. <b>Air Reading</b> – technical readiness audit. <b>IHE Partners</b> – faculty liaisons confirmed. <b>Stanford</b> – RCT design validation.
Pilot & Refinement (SY 2026–27)	<b>Milestone 2 (Dec 2026):</b> ≥80% scheduled minutes delivered; median fidelity ≥85%. College-student feasibility confirmed by mid-year report. Stanford produces <i>Arkansas Rural Implementation Brief v1.0</i> . <b>Milestone 3 (May 2027):</b> Tutoring fidelity ≥85%; ADE issues <i>Tutoring Toolkit v1.0</i> ; IHE partners & Air Reading: refined college student training & protocols	<b>ADE</b> – data collection and dashboard oversight. <b>Air Reading</b> – training and fidelity monitoring. <b>IHE Partners</b> – recruit and supervise college students. <b>Stanford</b> – formative evaluation.
Confirmatory RCT (SY 2027–28)	<b>Milestone 4 (Oct 2027):</b> ≥90% treatment schools launched on time with full tutor staffing (≥ 100 college students trained). <b>Milestone 5 (Jan 2028):</b> Mid-year fidelity ≥85%; college students meet quality thresholds comparable to paraeducators.	<b>ADE</b> – project governance and data pipeline. <b>Air Reading</b> – implementation and quality monitoring. <b>Stanford</b> – impact analysis. <b>IHE Partners</b> – continued tutor management.

Table 1. Detailed Timeline with Milestones		
	<p><b>Milestone 6 (June 2028):</b> Grade 1 Impact Report issued; Tutor-Type Comparison Brief published by Stanford (ADE-approved).</p>	
<p>Duration &amp; Workforce Study (SY 2028–29)</p>	<p><b>Milestone 7 (June 2029):</b> <i>Grade 2 Duration Report</i> (1-year vs 2-year vs control); <i>Tutor-Type Brief v2.0</i>; <i>Pipeline Interim Report</i> (<math>\geq 50\%</math> college tutors express interest in teaching).</p>	<p><b>Stanford</b> – data analysis and reporting. <b>ADE + Air Reading</b> – continued delivery and monitoring. <b>IHE Partners</b> – pipeline tracking.</p>
<p>Persistence &amp; Dissemination (SY 2029–30)</p>	<p><b>Milestone 8 (Dec 2029):</b> Final Evaluation Report submitted; Rural State Literacy Toolkit and College Tutor Playbook released open-access; Pipeline Final Report (<math>\geq 15\%</math> EPP enrollment conversion; <math>\geq 10\%</math> rural teacher hires within 3 years).</p>	<p><b>Stanford</b> – final analysis and reporting. <b>ADE + Air Reading</b> – toolkit publication and state briefings. <b>IHE Partners</b> – pipeline documentation.</p>

**Risk Mitigation Framework**

- Tutor Staffing: Risk of insufficient qualified tutors or high turnover.
  - **Mitigation**: ADE + IHE Partners. Recruit 6 months early; maintain 10% standby pool; offer retention stipends; conduct weekly monitoring.
- Data Integrity: Risk of missing or inconsistent tutoring data.
  - **Mitigation**: ADE + Air Reading. Automated sync validation ( $\geq 98\%$  accuracy); monthly cross-checks by ADE Data Team.
- Fidelity Drop: Risk of inconsistent delivery quality at scale.
  - **Mitigation**: Air Reading + Regional Coops. Weekly fidelity sampling; retraining for tutors  $< 80\%$ ; cooperative-level corrective action plans.
- Financial Overrun: Risk of cost inflation or duplicate spending.

- **Mitigation:** ADE + Air Reading. Quarterly financial reconciliation; 5% contingency fund; external audit by ADE Finance.

**B.3. Relevance and Demonstrated Commitment of Each Partner**

The project is anchored in a cross-sector consortium uniting Arkansas's statewide tutoring infrastructure, a nationally recognized evaluation partner, higher education institutions, and regional implementation capacity. Each partner brings non-overlapping expertise essential to implementing, evaluating, and sustaining the project at scale.

<b>Table 2. Core Partners and Demonstrated Commitment</b>	
<b>Partner</b>	<b>Relevance to Project Success</b>
Arkansas Department of Education (Lead Agency)	Only entity capable of statewide coordination across 100 rural schools. Existing relationships with fifteen regional cooperatives and rural districts enable rapid recruitment and sustained support. \$33M HIT initiative (Jan 2024–present) serving 100+ schools demonstrates capacity to manage statewide literacy interventions.
Participating Rural LEAs (100 schools)	Essential for testing external validity across diverse rural contexts. Schools spanning Delta, Ozarks, River Valley; remote to rural-fringe; small to large provide real-world variation needed to prove the model works beyond pilots. Enable workforce RCT by hosting both college students and experienced paraeducators.
Air Reading (Implementation Partner)	Providing WWC-standard evidence meeting standards without reservations (Neitzel & Storey, 2024). Arkansas track record ( [REDACTED] ) proves readiness for statewide expansion. Virtual platform uniquely addresses rural constraints: reaches remote schools, supports low-bandwidth environments, eliminates transportation barriers.
Arkansas IHE Partners	Existing EPP programs, field placement systems, and work-study infrastructure provide the only viable pipeline to recruit 100-150 college students annually.
Stanford University	[REDACTED] provides

Table 2. Core Partners and Demonstrated Commitment	
	<p>national credibility essential for findings to influence policy. NSSA advances research on high-impact tutoring cost-effectiveness and enabling conditions for implementation—directly aligned to this project's dual focus on literacy gains and workforce expansion. Led prior RCTs with embedded workforce comparisons—the exact design this project requires. Developed standards informing state and federal tutoring guidance nationwide.</p>

**Partnership Coordination and Sustainability:** Each partner fills a unique, essential role; no partner is redundant. Monthly cross-partner coordination meetings (chaired by [REDACTED]) and a shared project management platform ensure alignment and rapid problem-solving. Quarterly IHE partner meetings review college student recruitment, retention, and pipeline progress.

Each partner's formal commitment—supported by Commitment Letters, MOUs pending award of the project, resource allocation, and leadership engagement—ensures feasibility, accountability, and sustainability beyond the grant period. This consortium forms a durable statewide infrastructure for evidence-based literacy tutoring and workforce innovation that will persist after federal funding ends.

**B.4. Plan to Deliver Project Services More Efficiently at Scale and Maintain Effectiveness**

The project delivers high-quality, evidence-based tutoring at a statewide scale through a model designed for efficiency, consistency, and reach.

**Efficiency Mechanisms:** Air Reading's virtual platform tests statewide implementation with college student tutors. Virtual delivery eliminates travel costs and serves remote schools. Shared infrastructure (dashboards, rubrics, training) amortizes costs across 100 schools. Regional cooperatives provide efficient technical assistance. IHE partnerships leverage university systems (career services, work-study) for recruitment and supervision.

**Quality-Preserving Mechanisms:** Standardized structure maintains effectiveness. Tutoring is embedded in master schedules. Tutoring scope and routines mirror Tier 1 ELA for immediate skill transfer. All tutors receive identical training (8-hour Science of Reading core + monthly refreshers), follow standard rubrics, and use ADE-aligned frameworks. Coach ratios ( [REDACTED] ) enable proactive fidelity management. Fidelity monitoring ensures quality, as shown in *Table 3*.

<b>Table 3. Fidelity Monitoring</b>			
<b>Fidelity Dimension</b>	<b>Data Source</b>	<b>Threshold</b>	<b>Action Trigger</b>
Dosage (scheduled sessions delivered)	Platform logs; tutor session logs	≥80% delivered	<75% for 2 consecutive weeks → ADE site visit within 5 days
Adherence (following Science of Reading routines/coaching)	[REDACTED] [REDACTED]	Median [REDACTED]; bottom quartile [REDACTED]	[REDACTED] [REDACTED] [REDACTED]
Quality (pacing, modeling, feedback, student response opportunities)	[REDACTED] [REDACTED] [REDACTED]	Median quality rating [REDACTED]	[REDACTED] [REDACTED] [REDACTED]
Engagement (student attendance; teacher Collaborative Team meetings participation)	Platform attendance; Collaborative Team meetings sign-in	Student attendance ≥85%; teacher Collaborative Team meetings attendance ≥90%	Low attendance → family engagement outreach + scheduling adjustments
College Student Parity	Platform tutor-student parity reports	≥85%; no significant difference from experienced paraeducators	<85% → intensive coaching; gap → adjust protocols

**Cost-Effectiveness at Scale:** Virtual delivery and statewide infrastructure reduce fixed costs.

The dual-track workforce study quantifies whether college students achieve comparable outcomes at lower cost. The RCT ensures Arkansas scales only cost-effective versions.

**B.5. Quality of Mechanisms to Broadly Disseminate Information and Resources**

The project produces two comprehensive implementation guides enabling rural states to replicate evidence-based early literacy tutoring:

1. **Rural State Literacy Tutoring Toolkit** integrates findings on state adaptation protocols, rural implementation strategies, cost-effectiveness analyses, and policy alignment guidance.
2. **College Student Tutoring Workforce Playbook** provides complete protocols for IHE partnership structures, recruitment strategies, training sequences, supervision structures, retention strategies, pipeline conversion tracking, and cost models..

<b>Table 4. Dissemination Products Tailored to Audiences</b>		
<b>Audience</b>	<b>Key Products (High Utility)</b>	<b>Dissemination Channels (Broad Reach)</b>
Arkansas Educators & Leaders	Toolkit/Playbook components; PD modules	ADE website; cooperatives; AAEA conferences; Regional Collaborative Team networks
National Policymakers	Toolkit; Playbook; Return-on-Learning Brief (cost comparisons); WWC Report; legislative summaries; Pipeline Report (EPP enrollment, teacher hiring)	CCSSO (50 SEAs); NCSL (state legislatures); Results for America; NREA; ECS
Researchers	Peer-reviewed articles (workforce, cost-effectiveness, duration); pre-registered analysis plans; public-use data (FERPA-compliant)	AERA; SREE; WWC; journals (JREE, EEPA, AERA Open, Econ of Ed Review)

Table 4. Dissemination Products Tailored to Audiences		
Higher Education Institutions	Playbook; IHE case studies (Arkansas experiences); recruitment and retention briefs	AACTE (all EPP programs); ATE; Chronicle; EPP networks
Families & Communities	Summaries (English/Spanish); infographics; engagement resources	School communications; local media

**Timeline:** Toolkit (2028); Playbook (2029); Reports and publications (2029-2030)

**Active Support for Adaptation and Replication:** All products will be openly licensed (Creative Commons) and hosted on ADE, REL Southwest, WWC, and ERIC portals for free access. ADE, Stanford, and Air Reading will jointly provide:

**Technical Assistance:** Quarterly webinars (2029–30) for SEAs/LEAs on implementation, scheduling, college student workforce; direct TA to 3-5 rural states (site visits, data-sharing, IHE partnership guidance); Rural HIT Community of Practice (monthly virtual meetups with MT, WY, SD, ND, WV, MS, NM, ID, AK); IHE peer consultation (Arkansas IHE partners share recruitment, training, supervision, pipeline protocols with other universities).

**Peer Consultation Network:** Arkansas IHE partners share recruitment, training, supervision, and pipeline protocols directly with universities in other states seeking to implement college student tutoring models.

**C. Quality of Project Design**

**C.1. Logic Model: Quality of Conceptual Framework Linking Inputs to Outcomes**

Air Reading's theory of action: foundational literacy accelerates for Level 1 early-grade students when explicit instruction is delivered frequently in very small groups; when Tier 3 tutoring is tightly aligned to Tier 1 scope-and-sequence, routines, and assessments; and when delivery is reliable because time is protected, staff are coached, and data drive rapid adjustments.

A critical component is testing workforce innovation: Can college students receiving intensive training and supervision produce comparable literacy gains to experienced paraeducators? The theory posits that systematic training, ongoing coaching, and structural supports matter more than prior experience. If validated, this demonstrates that states can meet rural literacy goals without reliance on scarce certified staff—shifting scalability constraints from personnel to systems. This would dramatically expand rural workforce capacity and build teaching pipelines, potentially unlocking sustainable rural literacy solutions nationwide.

The logic model in Appendix G illustrates how inputs, activities, outputs, and outcomes connect in a disciplined, testable way: Careful research-driven design → high-quality implementation → students receiving personalized instruction from an engaging, motivating tutor → literacy gains at scale.

**Inputs - Resources and Capacity:** Table 5 outlines how project inputs, including state capacity, implementation partners, workforce, infrastructure, and enabling policy, combine to create conditions for reliable and high-quality delivery.

<b>Table 5. Project Inputs</b>		
<b>Input Category</b>	<b>Specific Resources</b>	<b>How Inputs Enable Outcomes</b>
State Capacity	ADE leadership; literacy/data teams; regional cooperatives	Statewide coordination, policy alignment, TA infrastructure; HIT initiative creates readiness
Implementation Partners	Air Reading platform (Science of Reading); Stanford evaluation (RCT design, WWC reporting)	Proven intervention + rigorous independent evaluation ensures quality and credibility of findings
Workforce	Experienced paraeducators + college students (diverse majors via IHE partnerships: EPP, STEM, liberal arts, business); coordinators;	Dual-track staffing addresses rural shortage; embedded RCT tests workforce viability; builds teaching pipeline

Table 5. Project Inputs		
	coaches; principals	
Shared Infrastructure	Training (8 hrs + monthly refreshers); observation rubrics; progress-monitoring (DIBELS 8/FastBridge); ATLAS ELA; data integrations; family toolkits	Standardized tools enable quality at scale; shared assessment for placement/pacing; identical training both tutor types ensures consistency
Enabling Policy	MOUs (100 districts); IHE partnership agreements (5–8 institutions); FERPA DSAs; stipend structures	Formalizes commitments; removes barriers; enables college student recruitment and pipeline tracking

**Activities - Operationalizing the Theory of Action:** Active ingredients (components hypothesized as critical to outcomes): (1) frequent, small-group tutoring aligned to Science of Reading; (2) explicit Tier 1↔Tier 3 coherence; (3) protected school-day time; (4) bi-weekly coaching cycles; (5) real-time data for adjustment; (6) intensive training/supervision enabling college students to produce comparable effects to experienced paraeducators.

Table 6. Project Activities		
Activity Cluster	Specific Actions	Link to Outcomes
Onboarding & System Build	By Aug 2026: Pilot complete (~20–25 schools); schedules locked, DSAs executed, integrations live. By Aug 2027: Full onboarding (50 treatment + 50 control)	Ensures infrastructure reliability before tutoring begins; protects against implementation failure
Training	[REDACTED]; identical for experienced paraeducators and college students	Builds tutor skill; common training ensures consistency; tests whether training equalizes performance
Tutoring Delivery	Grade 1: 4 sessions/week, 30 min, 1:2–1:3	Dosage + alignment = active ingredients;



**Outputs - Implementation Reach and Quality Signals:** *Table 7* outlines the project outputs, which signal whether implementation is reaching students at sufficient intensity and quality.

Table 7. Project Outputs		
Output Type	Specific Measures	Targets
Personnel Trained	Tutors, site coordinators, principals, Grade 1 teachers engaged in Collaborative Team meetings alignment	Pilot year: ~75–100 personnel trained Confirmatory year: ~150–200 tutors, 50 coordinators, 50 principals, ~100 teachers
Students Served	Level 1 first-graders receiving tutoring	Confirmatory year: ~750 students (50 treatment schools × 15 students/school)
Dosage Delivered	Sessions per student; scheduled minutes delivered; average group size	≥65 sessions/student annually; ≥80% scheduled minutes; avg group size ≤1:3
Quality Indicators	Bi-weekly scored observations per tutor; school dashboards updating weekly; district memos documenting progress	≥12–16 observations/tutor/year; 100% schools with live dashboards; monthly district memos
Family Engagement	Touchpoints per tutored student per term	≥2.5 touchpoints/student/term (reports, webinars, SMS) = ~2,000+ touchpoints annually

**Outcomes - Short-Term, Intermediate, and Long-Term:**

Short-Term Outcomes (by June 2027 – Pilot Year)

- System readiness: Protected tutoring blocks in master schedules
- Staff capacity: College students meet quality thresholds comparable to experienced paraeducators
- Infrastructure: Data integrations live; IHE partnerships operational
- Student skills: ≥15-point increase baseline to mid-year in phonics mastery
- Fidelity: Median adherence ≥85%; delivered dosage ≥80%

Intermediate Outcomes (by June 2028 – Confirmatory Year)

- Student impact: Primary outcome ATLAS Reading/ELA; intermediate DIBELS 8/FastBridge. Grade 1 ITT target  $\geq 0.17$  SD vs. control
- Workforce effectiveness: Students tutored by college students show gains comparable to experienced paraeducators (no significant difference)
- Equity: Positive point estimates for ELs, students with disabilities, economically disadvantaged students
- Practice change:  $\geq 75\%$  of Grade 1 teachers use aligned practices  $\geq 2\times/\text{week}$
- Reliable delivery:  $\geq 80\%$  of schools meet dosage and adherence thresholds

#### Long-Term Outcomes (by June 2029 and June 2030)

- Duration effects (2029): Randomized second-year test estimates  $\sim 0.25$  SD for two-year vs. one-year
- Persistence (2030): Grade 3 ATLAS ELA positive impacts;  $\geq 50\%$  of Grade 2 gains persist
- Sustainability: Districts codified playbooks;  $\geq 50\%$  of trained college tutors return or enter teaching pipelines;  $\geq 15\%$  EPP enrollment within 2 years;  $\geq 10\%$  hired as Arkansas rural teachers within 3 years
- Workforce pipeline: Pipeline conversion rates documented (EPP enrollment, teacher hiring); cost-effectiveness data enables informed workforce decisions

#### **C.2. Goals, Objectives, and Outcomes**

**Overarching Goal:** Establish a proven, scalable, sustainable model for delivering evidence-based foundational literacy instruction that significantly improves K-3 outcomes for Level 1 students in rural, high-need Arkansas schools and is replicable in states with similar rural populations.

**Specific Objectives with Measurable Targets:** Table 8 provides project objectives.

Table 8. Project Objectives			
Objective	Measurable Target	Deadline	Ambitious, Yet Achievable
1. System Readiness	Complete onboarding for ≥20 pilot schools (2026) and 50 treatment + 50 control schools (2027); establish 5–8 IHE partnerships with designated faculty liaisons	Aug 2026; Aug 2027	<u>Ambitious:</u> Coordinating 100 schools statewide simultaneously <u>Achievable:</u> ADE existing infrastructure (cooperatives, HIT initiative); phased approach allows learning
2. Confirmatory Grade 1 Impact	≥0.17 SD ITT effect on state ATLAS Reading/ELA vs. control group	June 2028	<u>Ambitious:</u> Replicating RCT effects at 100-school scale <u>Achievable:</u> Grounded in Johns Hopkins RCT; MDES calculation; pilot year stress-tests operations
3. Student Growth on Intermediate Outcomes	≥60% of tutored Level 1 students meet individualized growth targets on DIBELS/FastBridge (from ~45% control baseline)	June 2028	<u>Ambitious:</u> Serving lowest-performing students (Level 1) <u>Achievable:</u> Aligned dosage (≥65 sessions); small groups (≤1:3); Science of Reading routines
4. College Student Quality Parity	College students meet ≥85% quality thresholds on observations; students tutored by college students show literacy gains comparable to those tutored by experienced paraeducators (no significant difference)	June 2028	<u>Ambitious:</u> College students have no prior teaching experience; tests whether training/supervision can substitute for experience <u>Achievable:</u> Intensive supervision (██████); identical training; bi-weekly coaching; peer shadowing; proven curriculum; pilot validates feasibility
5. Implementati	≥80% of treatment schools meet both dosage (≥80%) and adherence	June 2028	<u>Ambitious:</u> Maintaining quality across 50 dispersed rural schools

<b>Table 8. Project Objectives</b>			
on Fidelity & Reliability	(≥85%) thresholds throughout the year		<u>Achievable</u> : Bi-weekly coaching; real-time dashboards; protected scheduling; pilot refinement
6. Teacher Practice Change	≥75% of Grade 1 teachers report ≥2×/week use of aligned practices; ≥0.5 SD growth on observation rubric	June 2028	<u>Ambitious</u> : Changing classroom practice, not just tutoring <u>Achievable</u> : Monthly Collaborative Team meetings; shared routines; visible student progress increases buy-in
7. Persistence to Grade 3	≥50% persistence of Grade 2 literacy gains on Grade 3 ATLAS ELA vs. control	June 2030	<u>Ambitious</u> : Effects often fade; testing 2+ years post-tutoring <u>Achievable</u> : Coherence design builds classroom capacity; teachers retain aligned practices
8. Workforce Pipeline	≥50% of trained college tutors return for second term or express interest in teaching; track EPP enrollment and hiring rates	June 2029	<u>Ambitious</u> : Converting tutors to teacher pipeline <u>Achievable</u> : Work-study integration; mentorship from certified teachers; positive student outcomes increase motivation

**D. Quality of Project Evaluation or Other Evidence-Building**

The evaluation follows a cohort for four school years: pilot year (2026–27) to build reliable implementation, confirmatory impact year (2027–28), experimental second-year extension for half the original treatment schools (2028–29), and final year capturing persistence on Grade 3 state Reading/ELA (2029–30). In parallel, we test whether experienced paraprofessionals or certified teachers produce larger gains in comparison to Arkansas college students when both are trained and coached identically. The result is robust evidence on effectiveness, practical

pathways for rural scale with staffing constraints, and clear return-on-learning estimates Arkansas can use for sustainable investment.

**D.1. Methods of Evaluation and What Works Clearinghouse Standards**

The evaluation is a multi-site, school-level RCT with intent-to-treat estimation and hierarchical models respecting nesting of students within schools. The design explicitly aligns to What Works Clearinghouse (WWC) Group Design Standards without reservations through the design factors outlined in *Table 9*.

<b>Table 9 Design Features Meeting WWC Standards</b>	
<b>WWC Requirement</b>	<b>How This Study Meets It</b>
Prospective random assignment	100 rural elementary schools recruited from 217 eligible statewide; randomized 50 to treatment, 50 to control using computerized algorithm with blocking by region/size before any intervention
Pre-registered analysis plan	Analysis plan submitted to public registry (e.g., Open Science Framework, AEA RCT Registry) before confirmatory year, specifying primary/secondary outcomes, analytic models, subgroups, sensitivity analyses
Baseline equivalence	Document equivalence on prior achievement (state ELA, benchmarks), demographics (FRL, EL, disability, race/ethnicity), school characteristics (size, locale, regional service cooperative)
Appropriate modeling of clustering	Hierarchical linear models (HLM) with school-level random intercepts; student-level covariates (prior achievement, demographics) increase precision
Intent-to-treat estimation	Analyze all students in assigned condition regardless of individual take-up; primary analyses compare all students in treatment schools vs. all students in control schools
Transparent attrition tracking	Document school and student attrition at each wave; conduct sensitivity analyses (e.g., bounding approaches) if attrition exceeds WWC thresholds; report all attrition patterns transparently

Table 9 Design Features Meeting WWC Standards	
Independent evaluator	<div style="background-color: black; width: 150px; height: 15px; margin-bottom: 5px;"></div> (Stanford University) lead evaluation; independent of ADE policy/operations and Air Reading implementation; nationally recognized expertise ensures credibility

**Sample and Power:**

Sample: 100 schools randomized 50:50; ~750 Level 1 first-graders in treatment schools (average 15 per school) in confirmatory year (2027–28).

Power analysis: Conservative assumptions—school-level ICC=0.10; two-sided  $\alpha=.05$ ; power=.80; covariate  $R^2=.40$  (from prior achievement, BOY benchmarks). Under these assumptions:

- 2027-28 Grade 1 confirmatory contrast (50 treatment vs. 50 control): MDES  $\approx 0.17$  SD
- 2028-29 contrasts (each treatment arm vs. control): MDES  $\approx 0.21$  SD; Two-Year vs. One-Year incremental: MDES  $\approx 0.25$  SD
- Embedded tutor-type experiment (~750 students; 375 per type; school blocking): MDES  $\approx 0.16$ – $0.17$  SD

These targets are policy-relevant effect sizes for early literacy interventions, sufficient to detect improvements moving students materially closer to grade-level performance. *Table 10* outlines the project’s multi-year structure.

Table 10. Multi-Year Project Structure			
Year	Design Element	Sample	Outcomes
2026-27 (Pilot)	~20–25 treatment schools pilot for operational stress-testing	Level 1 first-graders in pilot schools	Fidelity metrics; mid-year benchmark growth (signals "taking"); lessons learned inform refinements

<b>Table 10. Multi-Year Project Structure</b>			
2027-28 (Confirmatory)	All 50 treatment schools implement mature model; 50 control schools continue BAU	~750 Level 1 first-graders (treatment); matched control sample	Primary: State ATLAS Reading/ELA (EOY) Intermediate: DIBELS 8/FastBridge subtests (BOY/MOY/EOY)
2028-29 (Duration Test)	Second randomization: 25 treatment schools extend tutoring to Grade 2; 25 do not; 50 control continue BAU	Original cohort entering Grade 2	Grade 2 EOY: State ELA + benchmarks for all three groups (Two-Year, One-Year, Control)
2029-30 (Persistence)	Follow-up only (no active intervention)	Original cohort entering Grade 3	Grade 3 state ATLAS ELA (tests persistence)

**Embedded Workforce Experiment:** Within each tutoring year, students in treatment schools individually randomized to paraprofessionals or certified-teacher vs. Arkansas college-student tutors. Assignment blocked by school; both tutor types receive identical training, coaching, materials. This nested RCT produces head-to-head impact estimates under identical implementation conditions—critical for workforce planning.

**Outcomes Chosen for Policy Salience:**

- Primary: Arkansas state ATLAS Reading/ELA (end of Grade 1, Grade 2, Grade 3)—drives accountability, family confidence
- Intermediate: DIBELS 8 or FastBridge subtests (phonological awareness, decoding, oral reading fluency, comprehension)—provide sensitive measures of foundational skill growth; help educators interpret which strands drive overall gains

- Secondary: Attendance during reading block/tutoring; student engagement indicators; teacher-reported skill mastery

All confirmatory analyses are ITT, comparing all students in treatment schools to all students in control schools regardless of individual take-up, with HLM including school-level random intercepts and student-level covariates (prior achievement, demographics) to increase precision and meet WWC baseline equivalence standards.

**Why This Design Meets WWC Standards Without Reservations:** The study uses prospective school-level randomization (gold standard for school-based interventions); documents baseline equivalence; employs appropriate statistical models accounting for clustering; conducts ITT analyses preserving randomization integrity; uses independent evaluator with nationally recognized expertise; and commits to transparent reporting of all results, attrition, and sensitivity analyses through pre-registration and public dissemination.

## **D.2. Evaluation to Provide Guidance for Replication or Implementation**

A central evaluation purpose is producing guidance rural districts can use immediately, not just scholarly findings. The study documents:

### **Implementation Components for Replication:**

1. Scheduling decisions that protect tutoring time and reduce missed sessions (where to place tutoring relative to Tier 1 reading blocks; avoiding conflicts with services/meetings; arranging supervision so classroom continues uninterrupted)
2. Alignment mechanisms that ensure Tier 1↔Tier 3 coherence under Science of Reading (shared scope-and-sequence; synchronized decodables; common error-correction routines; unified progress monitoring)

3. Group sizes and session frequencies that fit rural staffing realities while maintaining intensity  
Level 1 students require
4. Assessment integration (how to use BOY/MOY/EOY benchmarks to form groups, pace  
instruction, adjust tutoring plans)
5. Data flows from tutor logs and observation notes to dashboards that keep principals/coaches  
informed for timely adjustments

Because staffing is the pivot of rural scale, we translate tutor-type findings into explicit staffing models. If trained paraprofessionals and certified teachers produce indistinguishable gains in comparison to college students with coaching, the playbook specifies training time, observation cycles, supervision structures allowing districts to recruit/deploy/retain college students at scale (including through Arkansas IHE partnerships). If trained paraprofessionals and certified teachers produce larger gains, the playbook specifies where these educators yield greatest return (e.g., students with largest skill gaps; coaching leadership positions) and how to combine both workforce types to reach coverage targets while retaining quality. Coupled with two-year randomization results, guidance helps districts decide: extend tutoring into Grade 2 for original cohort, offer Grade 1 tutoring to next cohort instead, or do both in staggered rollout.

**Replication Products:**

1. Replication Toolkit: Step-by-step implementation guides; training sequences for tutors; coaching cycles and rubrics; fidelity checklists; sample data dashboards
2. Rural Scheduling & Capacity Playbook: Bell schedules making space for tutoring without sacrificing core instruction; staffing templates showing how to cover tutoring with existing personnel plus targeted hires; cost calculators converting staffing choices into budget lines

3. Return-on-Learning Brief: Costs and effects in simple, comparative format so district boards and community stakeholders understand trade-offs and benefits

Dissemination through Arkansas's existing technical assistance networks (educational service cooperatives) and national channels helps other states with similar rural challenges adapt the model while preserving essential ingredients.

**D.3. Measuring Fidelity of Implementation and Linking to Outcomes**

Fidelity is both a precondition for impact and a lever for improvement. The evaluation measures four dimensions with clear thresholds, outlined in *Table 11*.

<b>Table 11. Plan for Measuring Fidelity of Implementation</b>			
<b>Fidelity Dimension</b>	<b>Measurement</b>	<b>Threshold for Acceptable Implementation</b>	<b>Data Source</b>
Dosage	Scheduled sessions delivered; session length	≥80% scheduled minutes delivered; ≥20 minutes/session	Platform logs; tutor session logs (weekly)
Adherence	Following Science of Reading routines; coaching guidance	Median [redacted] adherence across schools; bottom quartile [redacted]	[redacted] [redacted] [redacted]
Quality	Pacing, modeling, corrective feedback, opportunities to respond	Median quality rating [redacted]	[redacted] [redacted]
Engagement	Student attendance; teacher Collaborative Team meetings participation	Student attendance ≥85%; teacher Collaborative Team meetings ≥90%	Platform attendance logs; Collaborative Team meetings sign-in sheets

Data streams feed school-level dashboards principals/coaches use to spot delivery gaps, adherence slips, or classrooms where transfer to Tier 1 is weaker than expected. Biweekly coaching cycles use these data to plan targeted support and track whether changes take hold.

**Linking Fidelity to Outcomes:** Analyses explicitly link fidelity to outcomes:

- Dose-response relationships: Estimate how many sessions and what session lengths associate with meaningful gains for Level 1 students
- Adherence/quality to intermediate skills: Model relationship between adherence/quality and DIBELS/FastBridge skills so coaches know which practice elements to prioritize
- Minimum thresholds by tutor type: Test whether college students require different coaching intensity to achieve similar results
- Protected time and coaching routines: Examine whether schools consistently maintaining protected tutoring time and coaching produce larger impacts, helping districts set right non-negotiables

This fidelity-to-impact linkage is essential for replication: it tells practitioners exactly what must be present for the model to work and where there's room for local adaptation without loss of effect.

Rigorous fidelity work produces immediate and cumulative benefits: principals manage tutoring as an integral part of the instructional program; coaches target time where it matters; tutors receive timely, concrete feedback improving student experience; state receives clear description of implementation conditions under which it can expect reported impacts. Because fidelity measures are built into daily practice, they don't impose heavy additional burdens—they provide the information backbone making rural-scale tutoring feasible.

#### **D.4. Design for Valid, Reliable Information to Guide Replication**

The evaluation is structured to produce evidence Arkansas can trust and use, with validity for both causal inference and practical replication.

**Internal Validity:** Secured through prospective, school-level randomization; pre-registered analysis plans; intent-to-treat (ITT) estimation; appropriately clustered models; and close attrition monitoring/reporting to meet WWC standards without reservations. Baseline equivalence is demonstrated on prior achievement, demographics, and rural criteria so confirmatory contrasts are unambiguous.

**External Validity:** Strengthened by a multi-site design across 100 rural schools, ensuring results reflect a range of community sizes, staffing patterns, and transportation constraints characterizing the state. Not a hothouse study—captures operational realities (coverage, small staff, competing obligations) that determine whether HIT reaches students reliably.

**Equity Focus Through Disaggregated Reporting:** Subgroup analyses by English-learner status, disability status, economic disadvantage (FRL), and race/ethnicity. Report point estimates and confidence intervals for each subgroup, and test treatment-by-subgroup interactions. All subgroup effects are reported transparently, regardless of statistical significance, to support equity-focused replication.

#### **Practical Documentation Enabling Adoption:**

**Cost Model:** Disaggregates fixed costs (curriculum licensing, training development, data integrations, coaching tools) from variable costs (tutor compensation by type, ongoing coaching, technology, supervision/coverage), making economies of scale visible. For each model variant (1-year vs. 2-year; experienced paraeducators vs. college students), reports cost per student served, cost per 10 hours of tutoring, and cost per additional month of reading growth. Because

the design produces both impact and cost for each option on the same cohort under identical operating conditions, resulting return-on-learning profiles are decision-ready: superintendents see, side-by-side, which combinations deliver the largest gains per dollar.

Staffing Pathways Documentation: Recruitment timelines, training sequences, coaching loads, observation cycles, and supervision structures for experienced paraeducators and college students—allowing other districts to replicate the workforce model.

Operational Templates: Bell schedules, duty rosters, coverage plans rural principals can adapt; family communication scripts; metrics/dashboards districts can adopt for real-time monitoring.

**Cost-Effectiveness Study**: Required by EIR statute, the cost-effectiveness analysis uses administrative and readily available data. The accounting framework separates fixed and variable costs to show economies of scale; reports cost per student served for each model variant; and reports cost per additional month of learning (effect sizes converted to months using national norms) for each variant. Enables transparent comparisons—1-Year Experienced Paraeducator vs. 1-Year College Student vs. 2-Year models. This return-on-learning analysis makes allocation decisions transparent and evidence-based, critical for sustainable scale.

**Dissemination for Broad Impact**: Independent evaluation report presents methods and findings with WWC-required detail. The Rural State Literacy Tutoring Toolkit, the College Student Tutoring Workforce Playbook, and a return-on-learning brief translate findings into immediately deployable resources. Results will be shared with Arkansas leaders through established channels and with national audiences (CCSSO, NCSL, AERA, SREE, WWC, NREA) so other rural states can adapt while preserving core elements.

**Public commitment:** The final evaluation report will be made public, submitted to Education Resources Information Center (ERIC), cost data, and complete analytical code (including statistical models, effect size calculations, and cost-effectiveness analyses) will be released for secondary analysis.

Arkansas will know not only whether Air Reading HIT works for Level 1 students in rural schools, but also how to make it work at scale, what it costs, which staffing models produce the greatest return in which settings, and whether adding a second year meaningfully changes long-term outcomes. By following students to Grade 3, the state will know whether early gains persist where accountability systems measure them. By investing in a study that is both rigorous and practical, Arkansas positions itself to deliver faster improvements for its most vulnerable readers in ways communities can sustain.

## References

Forward Arkansas. (2024). *Rural Education and Literacy Equity Report*. Little Rock, AR.

Institute of Education Sciences (IES). (2022). *What Works Clearinghouse Procedures and Standards Handbook, Version 5.0*. Washington, DC: U.S. Department of Education.

Learning Heroes. (2023). *Parents 2023: Going Beyond the Headlines*. Washington, DC.

National Assessment of Educational Progress (NAEP). (2022). *Reading Assessment Results*. Washington, DC: National Center for Education Statistics.

Neitzel, A. J., & Storey, N. (2024). *Air Reading: A randomized evaluation of a virtual tutoring model*. Center for Research and Reform in Education, Johns Hopkins University.