

Project Skyword Literacy: Lifting Decoding and Fluency through Data-Guided Practice**A.SIGNIFICANCE**

Project Skyword Literacy *addresses the absolute priority through (a) scaling of an entrepreneurial, field-initiated innovation to improve student achievement for high-need students and (b) expanding evidence-based literacy instruction.* This priority is achieved through a comprehensive plan to disseminate and rigorously evaluate WordFlight, an online platform supported with moderate evidence (McMurray & Klein-Packard, 2024; Appendix J1, WordFlight RCT; Appendix J2, Evidence Framework Alignment) and anchored in established approaches to literacy (Vaughn et al., 2022; Appendix J3, References) that enhance group and individual student decoding. Skyword introduces an innovative approach by extending prior evidence-based decoding instruction with a unique composition of components—diagnostic placement that separates knowledge from automaticity, and state-led implementation supports—to serve grades 3–5 students who receive supplementary instruction due to low-achievement in reading, including high-need subgroups, at scale. Specifically, this project addresses persistent decoding and automaticity gaps by deploying WordFlight’s explicit, skill-placed decoding intervention across districts, and conducting a confirmatory evaluation of standardized outcome measures and state proficiency assessments using a high-quality cluster RCT (see Appendix G, Logic Model).

A consortium led by the Wyoming Department of Education (WDE) and consisting of Foundations in Learning (FIL; Developer of WordFlight), the Iowa Reading Research Center (IRRC), and the University of Iowa’s Center for Evaluation and Assessment possesses the requisite experience and expertise in reading intervention, research, and implementation. Due to WDE’s leadership, Skyword *addresses the competitive preference priority related to returning education to the states.* WordFlight’s suitability for individualized instruction of high-need

students (i.e., those who persistently struggle with reading after receiving conventionally effective instruction) will *further address the competitive priority related to expanding access to high-impact tutoring through targeted activities in the final years of the project.*

The purpose of Project Skyword is to scale and rigorously evaluate WordFlight for struggling readers in grades 3–5, improving decoding and automaticity at district scale while establishing a replicable, state-led, evidence-based approach that can be adopted by other systems. FIL has introduced WordFlight into 123 schools; of these, 69% are Title I. In Wyoming, Iowa, and many other rural states, many students continue to score below proficient on measures of reading (e.g., National Center for Education Statistics [NCES], 2025), underscoring the need to sustain early gains through fluency. Project Skyword addresses this need. Specific objectives include: (1) Implement a high-quality experimental evaluation with high-fidelity; (2) Improve student reading outcomes; (3) Build educator capacity to implement and sustain WordFlight; (4) Maintain and sustain use of WordFlight in schools over the duration of the funded period; (5) Conduct exploratory tests of train-the-trainer and (6) high-impact tutoring applications of WordFlight; and (7) Scale, share, and ensure quality of project reports. To accomplish these objectives, Project Skyword leverages the expertise of its partners and grant funding to *implement and scale an innovative, evidence-based approach: WordFlight.*

Bridging the Science of Reading and the Science of Learning

Despite decades of progress in the Science of Reading, too many students continue to struggle with automatic word recognition and fluency (Roembke et al., 2021). Explicit, direct instruction—a cornerstone of the Science of Reading—effectively teaches elemental skills of decoding (Dilgard et al., 2022). However, it is through rich and varied practice that students consolidate and generalize their learning to achieve fluency and comprehension (Apfelbaum et

al., 2019; Odegard & Gierka, 2025). While some children do not receive adequate instruction in their early years, many who reach benchmarks in foundational decoding skills fail to sustain progress in developing automaticity and fluency in more complex text (Chall & Jacobs, 2003).

Cognitive science offers insights into how students move from initial skill acquisition to automatic, fluent performance (Roediger & Butler, 2011). Reading, like all complex skills, requires activation, storage, and efficient retrieval of information from long-term memory (Sweller, 2016). Automatization benefits from enhancements to explicit instruction that integrate structured, systematic, variable practice designed to strengthen memory networks and facilitate effortless access to stored knowledge (Odegard & Gierka, 2025). While explicit instruction effectively builds elemental knowledge—such as letter-sound correspondence—it is not an efficient mechanism for developing automatic retrieval (Oslund et al., 2018). A systematic, structured approach to “rich and varied practice” is necessary to support students as they transition from effortful decoding to fluent reading (Reed et al., 2019). Just as direct instruction is essential for learning the alphabetic principle and letter-to-sound correspondences, many students need structured, evidence-based practice opportunities intentionally designed to promote automatic word recognition and fluency (McMurray & Klein-Packard, 2024).

WordFlight is an innovative platform that combines technology-based individualized practice, guidance for teacher-led instruction, and foundational principles from both the Science of Reading and the Science of Learning to create an evidence-based, effective literacy intervention. By applying research on memory and learning processes, WordFlight operationalizes a systematic, structured approach to diagnosis, instruction, and practice that is responsive to individual student needs. This model aligns with contemporary learning frameworks (Odegard & Gierke, 2025) that target automaticity and fluency by engaging

both explicit and implicit learning systems. The Varied Practice Model (VPM; Apfelbaum et al., 2019) underlying WordFlight builds on decades of research demonstrating that variation in tasks, content, and difficulty enhances retention, generalization, and retrieval (e.g., Apfelbaum et al., 2013; Rost & McMurray, 2009; Rovee-Collier & Dufault, 1991). Exposing learners to multiple exemplars, orthographic patterns in linguistic contexts, and graduated levels of difficulty promotes durable learning (Apfelbaum et al., 2019). Unlike traditional blocked practice that isolates skills, VPM emphasizes systematic variation and relevant contrasts—such as interleaving examples and highlighting similarities and differences across content. This approach has been shown to improve decoding (Apfelbaum et al., 2013), oral reading fluency (Reed et al., 2019), and vocabulary retention (Rosa et al., 2017). *Thus, the project modifies standard explicit instruction by integrating a structured, variable-practice engine and teacher-guided consolidation, an emerging component that requires further testing at scale in grades 3–5.*

WordFlight Components

WordFlight is a blended intervention that combines technology-based individualized practice with teacher-led instruction. WordFlight has distinctive components relative to typical instruction, including: (1) Online Diagnostic Assessments, (2) Personalized Online Practice Lessons with Progress Monitoring, (3) Teacher-Driven Instructional Tools, and (4) Comprehensive Instructor Preparation. Each of these tools iterate upon and synthesize accepted practices in reading instruction while extending both the services provided by digital instruction and its integration into schools (Vaughn et al., 2022). The combination of components further constitutes the project’s innovative integration to test combined effects beyond any single component. Although, learners at the elementary (i.e., < grade 5) and secondary levels receive separate scope and sequence, the fundamental components of WordFlight remain consistent

regardless of grade level. Foundations in Learning has security measures in place to protect against the loss of data and encrypts information transmitted over the Internet. Access to information stored on servers and in databases is limited and by permission only. FIL's management team continuously reviews and updates WordFlight's integrity and the security of personal information. These protections, along with automated rostering, make statewide implementation feasible in rural contexts with thin staffing while safeguarding student data.

Online Diagnostic Assessments. Developed with support from the Institute of Education Sciences (Apfelbaum et al., 2018), the WordFlight Diagnostic is an online assessment that distinguishes between students' decoding knowledge and their automatic use of that knowledge. The WordFlight Diagnostic shows strong psychometrics: test-retest reliability typically above $r = .80$ across decoding and automaticity subscales, and predictive validity with standardized measures (e.g., decoding WRMT $r = .63$). With just three online sessions of 15-20 minutes, the WordFlight diagnostic pinpoints individual profiles across grapheme-phoneme classes and distinguishes knowledge (decoding) from efficient use (composite of decoding and automaticity via masked tasks), giving educators actionable targets for intervention and progress monitoring. This diagnostic precision enables educators to target the most efficient instruction and practice for each learner. Diagnostic data specify the appropriate curriculum for each student and interim objective assessments along with ongoing student performance drive the adaptive curriculum within WordFlight, ensuring that instruction is matched to student need. Results from pre-, mid-, and post-intervention Diagnostic administrations provide evidence of individual growth and allow for meaningful comparisons across learner profiles. This enables efficient targeting at scale—a prerequisite for cost-effective expansion.

Personalized Online Practice with Progress Monitoring. WordFlight is a digital intervention targeting skills essential for reading fluency—phoneme-grapheme correspondence, morphemic and syllabic knowledge, decoding and encoding, and automatic word recognition (Appendix J4, Learning Principles in WordFlight Units). The intervention is designed to enable students to extract both the regularities and irregularities of the English symbol-sound system. The curriculum is divided into 6 Objectives, each consisting of 4 instructional units, with routine assessments preceding and following every four units of content (Appendix J5, Overview of WordFlight Progression). Each objective systematically targets linguistic contrasts in the context of varying language units, tasks, and difficulty. The online practice of WordFlight provides over 10,000 structured, interleaved practice opportunities with words and their sublexical language units embedded in sentences and passages. Lesson items include text, pictures, and audio components (Appendix J6, Sample WordFlight Tasks). These features benefit learners who have not generalized early decoding skills to fluent reading (Magnan & Ecalle, 2006; Reed et al., 2019).

Students proceed through the WordFlight curriculum in a systematic order, in which tasks, support and difficulty level of wordlists are adjusted by student performance. Elements such as animated avatars and token economies are employed to enhance student engagement (Appendix J6, Sample WordFlight Tasks). Through systematic variation across multiple linguistic contexts, the intervention is designed to support generalization and make learning stick (Raviv et al., 2022). Objectives conclude with an assessment consisting of 20 trials. Students are advised to complete 20-min sessions during daily reading instruction on internet-capable devices (e.g., laptops, desktops). Student performance of individual lessons is available for immediate review (Appendix J7, Lesson Dashboard). Information the Diagnostic (Appendix J8, Diagnostic

Dashboard) and Objective tests (Appendix J9, Objective Dashboard) tasks is also progressively tracked, presented on an accessible dashboard, and linked to email notifications describing student progress to teachers (Appendix J10, Sample Teacher Notifications).

Teacher-Driven Instructional Tools. In contrast to conventional online literacy systems, WordFlight supports teacher-directed instruction by generating specific lesson plans that target different levels of instructional need. Teachers have access to lesson plans for each unit and zone of instruction. Using these materials and suggestions, teachers provide direct instruction to reinforce concepts covered in the online modules, provide feedback, scaffold learning, and provide motivation and support. Importantly, students are given the opportunity to rehearse and elaborate on concepts to deepen their learning and internalize their skills. Sample materials from the WordFlight instructional platform, assessment results, small group teacher support, and the WordFlight scope and sequence appear in Appendix J11 (Sample Lesson Materials). Each unit includes a teacher overview and a menu of 10-minute small-group lessons (often two per day) with clear instructions, word lists, and printable student pages. Activities blend decoding, spelling and syllable routines, quick games, and connected-text reading with comprehension prompts. Lessons follow an explicit sequence—introduce the target, model, guide practice (often with whiteboards), then read and discuss a short passage to consolidate and apply. The format is flexible so instructors can expand, contract, or repeat sessions based on student pacing needs.

Comprehensive Professional Preparation. A critical component of WordFlight’s impact—and one compatible with implementation at scale—lies in its coaching model. Professional learning and coaching are designed around three principles: (1) Learning about Learning, wherein teachers build foundational knowledge of how students acquire, generalize, and automate reading skills (e.g., linking science of reading and learning principles to read-

alouds, Appendix J12 Sample PD Lesson in VPM); (2) Collaborative Problem Solving, wherein educators learn to analyze student data, compare learner profiles, and test instructional hypotheses through peer collaboration (e.g., placement with the Diagnostic, Appendix J13 Sample PD Lesson in Data Analysis) and (3) Reflection and Feedback: Using immediate feedback loops, teachers refine their understanding of student progress in decoding and fluency development while developing skills to support observational skills (e.g., explicit instruction; Appendix J14 Sample PD Lesson in Skill Building). Through these experiences, teachers not only gain practical strategies for intervention but also internalize the principles of learning that underlie fluent reading. As a result, WordFlight fosters sustainable capacity building in schools—enhancing teacher expertise while improving student outcomes.

Research Support for the Proposed Intervention

Based on an independent evaluation—and consistent with Mid-phase requirements—**WordFlight is supported by moderate evidence with a well-designed experimental study (Appendix J1), warranting refinement and scaling with rigorous cost-effectiveness analysis.** An overview of the study’s alignment with an evidence framework appears in Appendix J2. McMurray and Klein-Packard (2024) conducted a randomized experiment with student-level assignment within schools/grades and pretest-adjusted analyses. Outcomes were standardized and technically adequate (e.g., WRMT-III), procedures were administered uniformly, fidelity was indexed via platform-logged dosage, and baseline equivalence was demonstrated across key demographics. Attrition was documented and not predicted by group, and the business-as-usual (BAU) control had no access to WordFlight. The authors found a statistically significant, positive effect of treatment ($F(1,256)=9.84$, $p=.0019$, $\Delta=.099$) for students who completed more than two-thirds of the intervention (i.e., 16 units; per-protocol subgroup—the present trial will

report intent-to-treat with dosage examined exploratorily). Effects on WRMT Word Attack ($\Delta=.14$) and Word ID ($\Delta=.107$) were significant ($p \leq .05$).

Although the prior study was in grades 6–7, the instructional mechanism evaluated—decoding leading to automaticity and then fluency—is identical at grades 3–5, where many students are consolidating decoding and building automaticity. WordFlight’s components are grade-agnostic; students are placed by skill, not age, and materials scale in complexity. Using the same practice elements in grades 3–5 is therefore an appropriate extension.

Bringing WordFlight to Scale: Project Skyword

Project Skyword will extend WordFlight’s innovation through a coordinated, statewide effort led by the WDE in partnership with FIL, the IRRC, and the CEA. WDE will oversee statewide implementation in Wyoming, district recruitment, and policy alignment; IRRC will recruit schools in Iowa and train hourly personnel to monitor fidelity administer assessment, and observe BAU instruction; FIL will provide the adaptive technology platform, technical assistance, professional learning, coaching and ongoing data security, and recruit schools outside of Wyoming and Iowa; and CEA will manage the evaluation at a high level (e.g., analysis). Together, these partners will establish a data-guided instructional cycle in which educators use WordFlight diagnostics and dashboards to inform placement, guide instruction, and track progress. Together, the partners will conduct a rigorous evaluation of an innovative intervention designed to improve literacy outcomes for high-need students.

B. STRATEGY TO SCALE

Project Skyword’s scaling strategy is rooted in the broader need for evidence-based literacy instruction in Wyoming, Iowa, and other rural areas that have approached FIL about WordFlight (e.g., Oklahoma and Montana; see Appendix C, Letters of Support). The effort to

meet this demand is state-led, cohort-based, and optimized for replication. All partners have agreed to the terms of the partnership (Appendix C, Letters of Support). Guided by a comprehensive management plan, the consortium will expand WordFlight across participating districts using standardized onboarding. Scale occurs in staged cohorts so that capacity, procurement, and training can feasibly occur. Division of duties ensures timely completion of key project activities. Existing FIL coaching “playbooks” will further ease speed of onboarding. Quarterly reviews led by WDE will facilitate adjustments to the project.

b.1. Demand for Broader Implementation of the Project

The demand for WordFlight exceeds current capacity and is aligned to our proposed scale. Current users come from 123 schools in 29 states. These schools are distributed across locations that are rural (23%), urban (41%), suburban (26%) and town (11%). Of these 123 schools, 69% are eligible for Title1 funding. **Rural school districts face persistent challenges in providing resources to schools, retaining qualified staff, and offering effective reading instruction** (Johnson et al., 2021). Although rural schools often have lower enrollments (Showalter et al., 2019), issues with transportation and appropriate staffing often make it difficult to provide targeted reading instruction (Hassel & Dean, 2015). Discrepancies in reading skills between urban and rural students at kindergarten persist after graduation (Clarke, 2014). Resource issues have coalesced to stymie Science of Reading reforms (Guitierrez & Terrones, 2023)—issues that WordFlight’s unique instructional capabilities have the potential to address.

Wyoming is a rural state with literacy gaps that underscore the need for evidence-based literacy instruction. Despite decreases in the number of students scoring Below Basic on Wyoming’s State Assessment (WYTOPP) between 2021 (28.10%) and 2024 (23.9%), the proportion of students identified as proficient has remained relatively flat (WDE, 2024).

Wyoming's 2024 NAEP results, with 64% of students scoring at- or Below Basic, growth in students scoring Below Basic, and stagnancy in the Proficient and Advanced bands, are warning signs (NCES, 2025). These data point to the critical importance of sustaining evidence-aligned instruction and targeted supports throughout the language and literacy continuum to ensure that early progress translates into long-term readiness and achievement. The Wyoming Literacy Needs Assessment further indicated that teachers required access to methods for individualizing instruction for students and high quality, evidence-aligned instructional materials.

Iowa, in addition to encompassing many rural areas, has also struggled to improve literacy outcomes for its students. On the most recent administration of the Iowa Statewide Assessment of Student Progress (ISASP) in 2024, a quarter to a third of all students in Grades 3 to 11 scored below proficient (IDE, 2024). Recent NAEP (NCES, 2025) scores paint a starker picture, with 70% of Iowa 4th graders scoring at or below proficient in reading. A substantial number of students with disabilities and English learners scored below proficient. For example, at Grades 4 and 8, approximately 68% to 70% of students with disabilities and 74% to 78% of English learners scored below proficient. Moreover, teachers repeatedly encounter challenges to routinely monitor student progress and implement targeted decoding instruction with fidelity (Brownell et al., 2010; Shanahan et al., 2025; van den Bosch et al., 2017). Recent policy changes in Iowa related to the use of instruction aligned with the Science of Reading accentuate the demand for evidence-based literacy instruction throughout the state.

Other rural states have encountered similar issues in achieving literacy outcomes for their students. In Oklahoma, recent NAEP scores (2025) suggest 77% of 4th grade students read at or below proficient. As in other rural areas, issues in reading often stem from access to effective reading instruction (National Council on Teacher Quality, 2023). According to Montana's

comprehensive assessment system (MontCAS), fewer than half of Montana’s students read at or above proficient (Montana Office of Public Instruction, 2024). District leaders report the same structural constraints seen in Wyoming and Iowa: thin staffing, limited access to evidence-aligned decoding interventions, and uneven professional learning opportunities—especially outside metro areas. These conditions create demand for a skill-placed, progress-monitored platform that can be delivered within existing literacy blocks and supported remotely at scale.

Project Skyword addresses this demand.

The suite of WordFlight components allow districts to deliver targeted decoding instruction within existing literacy blocks, meeting rural capacity constraints. The WordFlight platform is conducive to both small group and individualized instruction; additionally, the modules provide resources for teachers. The platform—combined with standardized onboarding, PD, and targeted coaching, provided through Project Skyword—is therefore oriented to the needs of rural states that continue to improve the capacity of personnel to deliver targeted literacy instruction. FIL’s current pipeline, letters, and usage data indicate more qualified sites than they can presently serve. The Skyword project proposes to recruit 80-100 schools and approximately 5000 students in schools across Iowa, Wyoming, and other LEAs in states such as Oklahoma and Montana that have pledged to support the project. The leadership of WDE and letters of support from the IRRC and various LEAs (Appendix C) provide evidence of demand and partner readiness for the expansion and evaluation of WordFlight at the proposed scale.

b.2 Feasibility of the Management Plan

One WDE Project Director [REDACTED] and counterparts from the IRRC [REDACTED], FIL [REDACTED], and the CEA [REDACTED] will lead the project. The Co-PDs of Skyword are qualified to lead the project and include experts who bring dedication and experience to

implementation and evaluation (Appendix B, Resumes of Key Personnel). WDE has experience building and leading collaborative efforts and has the support of the Governor as well as several internal subdivisions (e.g., Wyoming MTSS Center, Head Start). Consortium leaders will meet remotely on a quarterly basis for the duration of the project, and will convene in Wyoming once per year for the first two years. *Time commitments for key personnel and other staff appear in accompanying budget narratives.*

Division of responsibilities is aligned to partner strengths (Appendix J15, Management Structure Diagram; Appendix J16, Objective-Level Responsibilities and Accountability Matrix). WDE leads statewide coordination, contracting, and district recruitment; ensures policy and calendar alignment; and runs quarterly reviews. FIL provides the WordFlight platform, rostering and single sign-on set-up, technical assistance, a help desk, professional learning, and coaching, while also recruiting sites outside Wyoming and Iowa. IRRC manages district onboarding in Iowa and supports onboarding elsewhere as needed; assists with onsite confirmatory assessment, and conducts fidelity observations. CEA designs and runs the independent evaluation, including preregistration, data flows, analysis, and the ingredients-based cost study. Each organization has a coordinator and the necessary staff. The distribution of management efforts, timelines, and funds for project progress are adequate and will ensure that the project achieves specific objectives. **In accordance with funding guidelines, the consortium has committed to provide 10% of funds provided in the form of personnel time and WordFlight licenses.**

Timelines and milestones are cohort based so that implementation is staged and manageable (Appendix J17, Project Timeline). Objectives are measurable. Examples include the schools onboarded per cohort, percentage of teachers completing professional development, and weekly WordFlight dosage. Budget discipline is maintained through reusable assets and a clear

division of labor. Each quarter, the team compares planned versus actual spending and workload; corrective actions can include rebalancing coach caseloads, consolidating training sessions, or adjusting cohort sizes within the approved scope.

b.3 The Relevance and Demonstrated Commitment of Partners

The consortium is composed of partners with relevant expertise and experience that contribute critical resources and are deeply committed to achieving project goals. The **Wyoming Department of Education (WDE)** enables governance and dissemination across a range of LEAs with a demonstrated need for the intervention. [REDACTED] will direct WDE efforts on Project Skyword. [REDACTED] has over 24 years of experience in instructional leadership and coaching at the classroom, school, district, state, and national levels, and has previously managed large-scale literacy projects within Wyoming. WDE will recruit Wyoming LEAs and regional education centers, align project timelines with approved literacy screening calendars, and issue FERPA-compliant templates and secure file transfer protocols to the evaluator. WDE will chair quarterly project management meetings and assessments of project performance. Quarterly meetings will assist in maintaining a log of issues and taking corrective actions. Wyoming will manage budgets related to school stipends and hourly personnel needed for implementation for the entire project, regardless of location. Schools will receive stipends to support cooperation with the project (e.g., coordination of assessments). Director [REDACTED] and the WDE coordinator will lead statewide communications and coordinate with consortium partners, contract with consortium partners, ensure compliance with federal and state requirements, and execute MOUs with LEAs with a delayed-start clause for waitlist schools.

Foundations in Learning (FIL) is the creator of WordFlight and has successfully equipped districts across Iowa and other rural states with both the platform and training needed

to yield improved student literacy outcomes. FIL Director [REDACTED] has over 30 years of experience in curriculum development and executing reading research in schools. Additional staff include IT experts, a project coordinator, and implementation coaches. FIL will provide and maintain the WordFlight platform and assist consortium partners in LEA onboarding and day-to-day implementation support with on-site staff and telecommunications support. As a business with connections across the U.S., FIL will assume primary responsibility for cohort recruitment outside of Wyoming and Iowa (e.g., Oklahoma, Montana). FIL will also provide the blueprint for ongoing coaching, coach-to-school-ratio targets, and professional development. Additional responsibilities include (a) license provisioning and automate rostering, (b) co-development and administration of initial training, booster sessions, and instructor remediation; (c) usage/fidelity dashboards for schools and LEAs (minutes, sessions, units, mastery); and (d) quarterly implementation briefs describing uptake, fidelity, progress on platform-aligned mastery and standardized measures, and technical integrity reports regarding sites who fall below dosage thresholds. FIL will be primarily responsible for administration of exploratory evaluation activities in Years 3-5. FIL will furnish reports of usage, dosage, and mastery meta-data for ongoing implementation monitoring.

The **Iowa Reading Research Center (IRRC)** is a mission-aligned state center housed at the University of Iowa with sustained legislative support (annual state appropriation of approximately \$1.5M) and statewide service to Iowa schools. The leadership team for Skyword—[REDACTED] (Director), [REDACTED], and [REDACTED]—brings complementary expertise in designing and evaluating literacy interventions for students with learning disabilities, research design, intervention implementation, and professional development. IRRC's commitment to Skyword is concrete and ongoing. The center will co-lead implementation with FIL by managing

Iowa district onboarding. They will also train hourly personnel in instructional-integrity monitoring, assessment administration; and BAU documentation using a standard observation protocol. The IRRC will leverage its extensive relationships to recruit Iowa schools for the project and maintain secure data transfer to the evaluator. IRRC will participate in quarterly governance meetings led by WDE, assist FIL in assessing fidelity and BAU, and assist in the creation of materials for dissemination (e.g., implementation guides). This capacity, infrastructure, and record of successful grant administration demonstrate both the relevance of IRRC to Skyword's goals and its commitment to delivering high-quality, scalable supports throughout the project period.

As an independent, nonpartisan research organization, the University of Iowa's **Center for Evaluation and Assessment (CEA)** is a team of faculty, staff, and graduate students who conduct multiple forms of program evaluation and assessment in collaboration with colleges, universities, and school systems around the world. The mission of the CEA is to provide high quality program evaluation and research services to support data-driven decision making. The CEA has completed 150+ funded evaluations using mixed methods designs. CEA Director [REDACTED] will oversee the evaluation team and CEA activities. [REDACTED] background in literacy education makes her uniquely qualified to lead an evaluation team for this project. Under [REDACTED], the CEA will complete preregistration of the statistical analysis plan prior to randomization; refine training protocols and test administration guidelines in Year 2; create CONSORT-flow charts; effect estimates, and the project data dashboard; and complete the cost estimate. Additionally, the CEA will prepare de-identified analysis code and replication materials at the conclusion of the project.

b.4 Quality of the Plan to Deliver Efficient, Effective Services at Scale

Skyword will deliver services efficiently at scale while maintaining effectiveness through a disciplined operating model. Efficiency is achieved through delegation of tasks to appropriate consortium partners; reusable, remote/online avenues of PD and service delivery; CEA's data dashboard for the dissemination of relevant information to stakeholders; and a management structure emphasizing objective measures of efficiency (e.g., onboarding time, coach hours per school, per-student delivery cost) throughout implementation. FIL has experience leveraging telecommunications for WordFlight's implementation and will collaborate with the consortium to ensure coordinators receive assistance needed over the course of the project. The division of labor between CEA and IRRC will ensure an independent, comprehensive evaluation and prompt attention to in situ logistics. The coach-to-school ratio will improve over the course of Years 1-2 as playbooks and exemplars reduce touch time. Professional learning combines short asynchronous modules with live sessions, reducing seat time while preserving practice quality. Effectiveness is safeguarded by dosage and fidelity thresholds (target 60–80 minutes per student per week; warnings given if below target for two consecutive weeks) and weekly dashboards with automated alerts for coach follow-up. Quarterly reviews led by WDE will foreground specific targets to guide project activities and improve services. In addition to monitoring and pursuing efficiency over the course of the project, we will conduct an ingredients-based cost study to document resources required to implement WordFlight at scale.

b.5 Quality of Mechanisms for Broad Dissemination

Skyword will make its methods and materials broadly usable through state-led and open channels. WDE will host quarterly statewide webinars and share practitioner briefs through SEA listservs and regional service centers. WordFlight is a licensed, proprietary platform (Appendix E); however, all implementation artifacts (e.g., new PD materials) and all evaluation artifacts

(e.g., preregistration, analysis code, de-identified dictionaries) will be disseminated openly. FIL/IRRC will publish an implementation toolkit (e.g., onboarding checklists, PD slide decks and micro-modules, fidelity rubric) and a replication pack for LEAs (e.g., cost planning worksheet). CEA will release accessible evaluation products—a preregistration link, interim memos, a final technical report, and a practitioner summary—and, data dictionaries, and analysis specs in an open repository so other entities can adapt the approach. To support adaptation and replication, the consortium will offer office hours each term, maintain a public FAQ, and identify mentor schools from Cohort 1 to advise new adopters.

C. QUALITY OF THE PROJECT DESIGN

c.1 Quality of Logic Model

The logic model (Appendix G) specifies a coherent, evidence-aligned pathway from state-level inputs to student outcomes. Inputs include WordFlight licenses; state and regional implementation capacity (e.g., WDE state level coordination, FIL PD and coaching); independent evaluation (CEA); teacher time and devices (LEAs); and standardized assessments (e.g., Word Attack, Reading Fluency, state ELA). WordFlight contributes diagnostic assessment, adaptive scheduling of decoding practice, and real-time assessment often absent from BAU, providing a direct pathway from explicit instruction to measurable gains. Additional resources fund activities that are directly tied to the intervention’s mechanism: diagnostic placement; explicit, cumulative decoding instruction; teacher professional learning (asynchronous modules plus live sessions); coaching; and data dashboards for continuous feedback. Preconditions include completing rostering and PD before the baseline window. The model translates these into verifiable outputs—schools onboarded, teachers trained, percent of classrooms meeting fidelity, percent of students meeting weekly dosage, and on-time data deliveries—which serve as

proximal indicators that planned work occurred with quality. For control/waitlist settings, BAU instruction is documented with the ICE-R protocol to confirm no WordFlight exposure and to characterize reading-block content. This strengthens the link between WordFlight and observed outcomes and identifies potential mediators or barriers to program success.

The causal chain is explicit and testable. Dosage and instructional fidelity are hypothesized mediators. Dashboard flags trigger coach outreach and targeted PD adjustments. In the short term, students' gain in decoding and automaticity, reflected on standardized measures such as the Word Attack subtest of the WRMT-III, and the TOSCRF. These learning changes, coupled with teacher adherence to explicit routines, yield medium-term action changes: higher classroom fidelity, sustained student practice, and increased rates of students meeting grade-level ELA proficiency. Over time, this scales to long-term conditions: durable reading proficiency, reduced remediation needs, and a practical pathway for targeted tutoring for non-responders. The model incorporates equity monitoring by disaggregating dosage, fidelity, and outcomes for ELL and IEP subgroups and by setting gap-closing targets. Across cohorts, the project targets a meaningful reduction in dosage gaps for ELL and IEP subgroups.

The framework also embeds feedback and governance: usage and fidelity dashboards trigger coaching outreach and PD adjustments, and quarterly reviews track acceptance criteria (e.g., dosage) without changing core components and compare cohort outcomes to prior cohorts to verify maintained effects during later project phases. Efficiency indicators (e.g., coaching hours per school) are tracked alongside outcomes to ensure gains are maintained as supports become more efficient. Assumptions and risks (e.g., staffing) are identified with mitigations (WDE escalation). Finally, the logic model aligns to the evaluation: CEA will preregister confirmatory outcomes and analyze pretest-adjusted effects under a cluster RCT, while also

examining the hypothesized mediators (dosage and fidelity) that connect activities to outcomes. Together, these features demonstrate a high-quality conceptual framework in which each input is plausibly and measurably related to the intended outcomes.

c.2. Alignment of Measurable and Achievable Goals with Program Purpose

The objectives driving Project Skyword are measurable, achievable, and **clearly aligned with the purpose of expanding evidence-based literacy instruction**. The confirmatory goals (Years 1-2) of Project Skyword are to determine the impact of WordFlight on (1) decoding (WRMT-Word Attack, Woodcock, 2011), (2) silent reading fluency (TOSCRF-2, Hammill & Lee, 2014) and (3) the proportion of students that meet or exceed state reading proficiency, relative to BAU. These outcomes are pre-specified, measurable at the level of students targeted in Years 1-2 (grades 3-5) and collected in the spring and fall of every project year. The analysis further permits the evaluation of treatment across subgroups (e.g., ELL, special education) and other moderators of WordFlight's effectiveness. The confirmatory phase of the project (Years 1-2) will produce evidence designed to meet the WWC's standards, without reservations (see Section D, Quality of the Evaluation). Continued implementation of WordFlight in schools (Years 3-5) within the original two cohorts will provide insight into the maintenance.

In Years 3-5, the project will extend learning through exploratory evaluations of a train-the-trainer model, wherein onsite personnel will assume all on-site responsibilities associated with launching WordFlight, and high-impact tutoring for students with longstanding reading difficulties in the late elementary grades and middle school. For train-the-trainer, the objective is to build district capacity to deliver WordFlight PD and coaching with quality comparable to consortium delivery, evidenced by facilitator completion and readiness checks, teacher participation with brief post-assessments, and classroom fidelity scores. Community members,

such as volunteer faculty from the University of Wyoming, will participate training at no additional cost as an additional means of extending outcomes beyond the life of the project (Appendix C, Letters of Support). The tutoring element intentionally includes students outside of the grade band selected for the confirmatory study, as evidence suggests these students require more intensive, high impact reading instruction to achieve gains (Lovett et al., 2012). The objective is to provide targeted decoding and automaticity practice to students who remain below progress following extensive exposure to BAU intervention, with adherence to a simple session protocol, documented attendance, and monitoring of pre-to-post gains on the same standardized outcomes used in the confirmatory phase. Both strands include feasibility objectives—on-time scheduling, materials uptake, and maintenance of core classroom implementation—so that results can guide statewide adoption without increasing unit costs or diluting effects.

Implementation objectives that enable primary literacy outcomes are likewise measurable, and includes schools onboarded by cohort and term, PD completion and readiness checks, dosage and fidelity thresholds (see Section D, Quality of the Evaluation), and on-time data deliveries. Project Skyword aims to scale WordFlight across Wyoming and other LEA partners with a cluster randomized evaluation. Achievability is ensured by standardized onboarding, a cohort rollout, and a clear division of duties within the consortium. The CEA's broader evaluation will also provide insight into critical issues in scaling the program, including factors that support and hinder implementation, sustainability, the elements of each program that are necessary to produce the intended outcomes; and any mid-course corrections that are required to improve the quality of the programs and increase the likelihood of meeting and sustaining intended outcomes. Additional information about CEA's evaluation appears in Appendix J18 (Evaluation and Data Management Summary).

In addition to student-outcome goals, the project includes a measurable efficiency objective aligned with the program’s purpose. Specifically, we will conduct an ingredients-based cost study to document total and per-student costs of implementing WordFlight at scale, and track unit-cost trends alongside decoding and fluency outcomes. The CEA will specify cost data elements, collection timing, and reporting and will produce term and annual profiles so decision-makers can judge whether effects are maintained as supports become more efficient. This objective is achievable within the grant period using standard logs, rosters, and invoices and will inform replication by other districts and states.

D. QUALITY OF THE PROJECT EVALUATION

The evaluation comprises confirmatory and exploratory studies. The confirmatory study will assess WordFlight’s impact on standardized measures in accordance with WWC evidence guidelines. Exploratory analyses will assess moderators, generalization, and implementation train-the-trainer, and high-impact tutoring. In addition to managing these analyses, the CEA will document stakeholder perceptions (Appendix J18, Evaluation and Data Management Summary).

D.1 Evidence of Effectiveness Meets What Works Clearinghouse Standards

Primary Confirmatory Study

The proposed confirmatory study will employ a two-arm, parallel cluster RCT, widely recognized as the most rigorous methodological approach for establishing causal inference in education research (Spybrook et al., 2020), in Years 1-2 of the project. Confirmatory outcomes are decoding, silent reading fluency, and performance on state ELA assessments. Power analyses suggest the proposed study is capable of identifying anticipated effects of WordFlight (Appendix J19, Statistical Analysis Plan and Power Analysis). Respective standardized measures will include the Word Attack Subtest of the WRMT-III Test of Silent Contextual Reading Fluency—

2 (TOSCRF-2, Hammill & Lee, 2014). The WRMT-III is an individually administered test of general reading ability. The Word Attack subtest assesses the ability to decode unfamiliar nonsense words using decoding and is well-aligned with the targets of the intervention. Test-retest reliability for word attack is excellent ($>.83$) and concurrent validity with related assessments (e.g. Test of Word-Reading Efficiency-Second Edition [TOWRE-2], Torgesen et al., 2012) is high (.60, Woodcock, 2011). The TOSCRF-2 (Hammill & Lee, 2014) is a brief, standardized, norm-referenced measure of silent contextual reading fluency. It can be administered individually or in small groups and yields both age- and grade-based norms. Performance reflects the speed and accuracy with which examinees process connected text during silent reading. Alternative form and test-retest coefficients are acceptable (0.82–0.93; Wissinger et al., 2024). Evidence of concurrent validity with commonly used reading measures is high (0.67; e.g., TOWRE-2, Torgesen et al., 2012). Familywise Type I error across the three confirmatory outcomes will be controlled using the Holm method; subgroup, mediation, and any additional outcomes are exploratory. State reading proficiency scores will be modeled with a pretest-adjusted logistic mixed-effects model using the same random and fixed effects structure.

The unit of randomization will be the school, with all eligible third- through fifth-grade classrooms within each school assigned to the same study condition. Random assignment at this level ensures equivalence between treatment and control groups at baseline on both observed and unobserved characteristics, thereby providing a valid counterfactual for estimating the effects of WordFlight relative to BAU instruction. A school-level cluster design is appropriate for both methodological and practical reasons. WordFlight is adopted and implemented at the school level, with professional development and coaching delivered to teachers across full grade cohorts. Randomizing within schools (by student or grade) would create risks of contamination.

In addition, student outcomes are correlated within schools (intraclass correlation), requiring analytic models that account for the nested data structure (Murray, 1998). Aligning the unit of randomization with the unit of adoption both preserves internal validity and enhances external relevance, since adoption decisions are generally made at the school or district level.

Schools will be randomized in a 1:1 allocation ratio to either treatment condition or waitlist control condition. Treatment schools implement WordFlight immediately (Year 1), with teachers trained and supported to deliver the program. Schools in the waitlist condition continue with current literacy practices during Year 1, then adopt WordFlight in Year 2. This structure balances the need for rigorous causal evidence with practical and ethical considerations, ensuring that all participants receive the intervention. The confirmatory intent-to-treat contrast uses spring Year 1 baseline and spring Year 2 outcomes with the waitlist serving as the contemporaneous control; analyses in Years 2 and 3 will be labeled exploratory for persistence and replication.

Analyses will follow a preregistered statistical analysis plan and will use pretest-adjusted mixed-effects models with random effects for school and cluster-robust standard errors. For the WRMT and TOSCRF, models include fixed effects for treatment, baseline score, and grade, with a random intercept for school and cluster-robust standard errors at the school level. State reading proficiency will use a pretest-adjusted logistic mixed-effects model with fixed effects for treatment, prior-year ELA status, grade, and state, and a random school intercept; testing windows and proficiency banding will be harmonized. Preregistered reports will describe inputs including: clusters per arm, average students per school, pretests with R^2 with outcomes, desired power, and the resulting minimum detectable outcomes. Primary analyses use an intent-to-treat population; exploratory complier-average causal effect estimates will instrument treatment uptake with randomized assignment. Pretests will be same-domain measures and strata fixed

effects will be included to reflect the blocked randomization. The primary confirmatory analysis will occur in Years 1-2, estimating the intent-to-treat effect of WordFlight by comparing outcomes of treatment schools to waitlist schools. This provides the rigorous counterfactual necessary for WWC standards "without reservations." In September of Year 2, when waitlist schools adopt WordFlight, the study will address two exploratory questions: (1) persistence of impacts in early adopter schools, and (2) replication among new implementers. Because there will be no contemporaneous control group in Year 2, these analyses will be descriptive rather than causal, but will provide important evidence on sustainability and scalability.

Randomization procedures will ensure integrity and transparency. Prior to assignment, schools will be stratified on key covariates to achieve balance between conditions. Planned stratification variables include (a) school need level (e.g., percentage of students eligible for free/reduced-price lunch or Title I status, or baseline performance level), and (b) locale/region (urban, suburban, rural, or state/LEA). Within each stratum, schools will be randomized using a computer-generated random sequence by the CEA. Randomization will occur after baseline data collection is complete, but before treatment assignment is revealed to participating schools, teachers, or data collection staff. This sequence ensures that baseline measurement is not influenced by knowledge of assignment and that any differential attrition between conditions can be accounted for in analyses. CEA will implement blocked randomization via reproducible code with allocation concealed until baseline files are locked. The statistical analysis plan and data plan will be preregistered on an open registry before baseline collection and any randomization, and the randomization seed will be archived. IRB approval and FERPA-compliant data-use agreements will be executed prior to baseline data collection.

Assessment windows are prespecified in the fall and spring (see Appendix J17, Project Timeline) with identical make-up rules across arms. For each cohort, data collection includes fall baseline in the entry year and spring outcomes in that same entry year. In Year 3, assessments will be collected in spring 2028 to evaluate persistence patterns among early implementers and replication patterns among new implementers. For treatment schools (Cohort 1), fall Year 2 data (approximately September/October 2026) will also be collected to assess maintenance of effects over the summer. All eligible students in grades 3–5 within participating schools will be included in the study sample. Test administrators and scorers will be masked to condition; standardized procedures and identical make-up rules will be applied across arms. BAU classrooms will be documented with a standard protocol (see d.3, Quality of Plan for Measuring Implementation Fidelity) to verify students have no WordFlight exposure and to characterize core reading instruction and alternative interventions for descriptive and sensitivity analyses. Potential spillover (e.g., staff moving between schools) will be monitored and affected sites for predefined sensitivity analyses will be identified. Assessors will be trained and certified. All accommodations will follow publisher and state guidance, with identical assessment windows and make-up procedures maintained across conditions.

The confirmatory study is designed to meet WWC standards without reservations, as it incorporates (a) random assignment at the cluster level, (b) procedures for establishing and adjusting for baseline equivalence (see Appendix J19 Statistical Analysis Plan and Power Analysis), and (c) monitoring of attrition at both the school and student levels to ensure compliance with WWC thresholds (WWC, 2022). We will report baseline equivalence tables by arm; if any standardized mean difference exceeds 0.05 the model will include that baseline covariate, and if any exceed 0.25 results will be labeled non-equivalent and interpreted per

WWC guidance. Primary analyses will use complete cases with documented make-ups; sensitivity checks will apply multiple imputation and inverse-probability weighting while preserving intent-to-treat. Reports will include a CONSORT-style chart showing schools and students randomized, excluded, analyzed, and analyzed by arm. Data sharing will follow FERPA-compliant agreements with encrypted transfer and de-identified analysis files provided to the evaluator. Overall, this two-arm, school-level cluster RCT with waitlist crossover maximizes internal validity, ensures alignment with WWC standards, and provides rigorous evidence about the effectiveness and scalability of WordFlight. We will release a reproducibility bundle (e.g., clean code) and conduct a blinded code review prior to unmasking arm labels.

Exploratory Studies

In addition to the primary confirmatory analysis, we will examine the viability of “training the trainer” models and the use of WordFlight for high-impact tutoring in Years 3-5. All exploratory analyses will be preregistered separately with their own analysis families and error-rate control. First, a train-the-trainer study will test whether district-based facilitators can deliver WordFlight professional learning and coaching with quality comparable to consortium-led delivery. Sites will be recruited in Years 3-4 and matched to contemporaneous consortium-supported sites on grade band and prior usage. Implementation evidence will include facilitator completion and readiness checks, PD attendance and coaching logs. Cost logs will capture facilitator time and materials to inform scale decisions. All analyses will be labeled exploratory and preregistered separately from the confirmatory contrasts.

Second, a **high-impact tutoring study consistent with the competitive preference priority of expanding education choice** will estimate the incremental benefit of scheduled WordFlight tutoring for students in cooperating middle schools recruited in Years 3-4. Within

participating schools, eligible students will be randomly assigned at the student level to tutoring or BAU supports, with intent-to-treat analyses using pretest-adjusted mixed-effects models (students nested in classrooms/schools). Randomization will be blocked by school and grade. Outcomes will mirror the confirmatory study prior to outcome analysis. Tutoring fidelity will be monitored with a brief session checklist (adherence to placement, minutes completed) and attendance from platform logs; make-ups and uniform testing windows will support reliability, and masking will be used where feasible. These randomized tutoring results will be reported as exploratory impact estimates, accompanied by feasibility, staffing, and ingredients-based cost profiles so districts can judge practicality alongside effect sizes.

d.2 Evaluation Generates Effective Replicability to Other Settings

The evaluation is designed to yield actionable guidance on strategies that can be replicated or tested in other settings. By pairing a preregistered cluster randomized design with standardized outcomes and uniform procedures, the study will produce valid, reliable impact estimates while documenting the implementation conditions under which effects emerge. Reports will present core components versus adaptable features, link dosage and instructional-integrity indicators to outcomes, and summarize heterogeneity by grade and student subgroups. BAU instruction will be characterized with ICE-R so adopters can judge fit, and sensitivity analyses will show how robust results are to real-world variations. To support testing and scale, the project will release replication materials—implementation playbooks, PD assets, data dictionaries, and de-identified code—alongside brief cost and staffing profiles, enabling districts and states to reproduce the approach or adapt it to comparable contexts.

d.3 Quality of Plan for Measuring Implementation Fidelity

Skyword will measure and use fidelity systematically, with CEA leading the evaluation and FIL/IRRC collecting operational evidence. Fidelity has three components: (1) Dosage, defined by platform logs as minutes of active WordFlight use; the threshold is 60–80 minutes per student per week, with flags after two consecutive weeks below target; (2) Instructional integrity, captured with a brief rubric during early coaching cycles (explicit modeling, cumulative review, corrective feedback); (3) PD readiness, verified by attendance and short checks for understanding (target $\geq 80\%$ correct). Fidelity data is instantly available to FIL and flows to CEA’s dashboards weekly. The IRRC will trigger outreach within five business days when a flag occurs, as school stipends support both data-coordinators and personnel responsibility for fidelity outreach. The WDE will review fidelity data quarterly and revise adherence plans as needed. The CEA will also track school- and student-level attrition, disaggregated by treatment group on a rolling basis using roster files. Differential attrition will be computed at both school and student levels. If projected rates approach WWC thresholds, the IRRC/FIL will initiate make-ups or records reconciliation to reduce loss. Confirmatory analyses remain intent-to-treat; if thresholds are exceeded, the preregistered analysis plan will specify sensitivity checks (inverse-probability weighting/multiple imputation and baseline-equivalence adjustments) and documentation of reasons for attrition. Dosage and instructional-integrity indicators will be related to WRMT–Word Attack and TOSCRF in preregistered exploratory models.

For BAU verification and context, CEA will use the Instructional Content Emphasis—Revised (ICE-R; Edmonds & Briggs, 2003) observation protocol to document core reading instruction prior to implementation in treatment and control settings—confirming no WordFlight exposure, characterizing emphasis on decoding/fluency, and providing descriptive covariates for sensitivity analyses. CEA’s analyses will link fidelity to outcomes. Additionally pretest-adjusted

models relate dosage and rubric scores (and ICE-R descriptors for BAU) to standardized decoding outcomes (Word Attack, Word ID, ORF) and state proficiency indicators, with subgroup checks (EL/IEP). Sites that fail to meet thresholds are analyzed separately (and in sensitivity analyses) to confirm that implementation quality predicts outcomes and to guide mid-course supports without compromising the confirmatory test.

d.4 Evaluation Generates Valid Evidence for Replicating Project Strategies

Skyword's implementation and evaluation are structured to generate usable guidance for replication while producing valid and reliable evidence of effectiveness. The design specifies what to implement (e.g., WordFlight; PD plus coaching), how to implement it (e.g., standardized onboarding, dosage thresholds), and how to verify BAU conditions (e.g., ICE-R observations). Effectiveness will be estimated with a preregistered, cluster randomized trial using pretest-adjusted mixed-effects models and standardized outcomes with assessor training, uniform test windows, and make-ups to support reliability. To support adaptation, the CEA will report implementation parameters alongside impact estimates, including dosage and instructional-integrity distributions, subgroup results (e.g., ELL, IEP), and sensitivity analyses. Replication materials— including facilitator guides, PD decks and modules, and a data dictionary with de-identified analysis code—will be released in accessible formats. Quarterly briefs will translate findings into actionable steps (e.g., onboarding sequence, common pitfalls, and remediation routines), and the final report will include a CONSORT-style flow and a section that maps required capacities for districts seeking to replicate or scale the approach. Cost analyses engendered by the CEA's ingredients-based cost study will further inform replication efforts.