TABLE OF CONTENTS

Preparing High-Need Students for Success in Early Science Instruction: An Early Phase Project

A. SIGNIFICANCE

This Early Phase research project aims to address **Absolute Priority 1** (Demonstrates a Rationale) and **Absolute Priority 3** (Field Initiated Innovations – Promoting Equity in Student Access in Educational Resources and Opportunities: STEM) by developing a kindergarten science program to empower students from historically underserved communities to explore and excel in the world of science. Specifically, the Scientific Scouts-Kindergarten (Sci-K) program will infuse academic language and early number sense skills with the kindergarten Next Generation Science Standards (NGSS Lead States, 2013) to promote equitable access to high-quality learning opportunities in early science. Our motivation to develop the Sci-K program is grounded in four premises. 1. Many U.S. students from underserved communities face structural inequities in science learning that are noted to begin as early as kindergarten (Morgan et al., 2016). 2. Findings from the 2019 National Assessment of Educational Progress (NAEP) Science survey indicate U.S. classrooms offer few opportunities for students to engage in scientific-inquiry related investigations and discourse (National Center for Educational Statistics [NCES], 2020). 3. There is a scarcity of evidence-based science programs tailored for kindergarten classrooms, which hinders the provision of equitable learning opportunities (What Works Clearinghouse [WWC], 2023). 4. It is imperative that kindergarten students from marginalized communities are granted equitable access to meaningful, engaging science instruction (National Science Teachers Association [NSTA], 2014; National Academies of Sciences, Engineering, and Medicine [NASEM], 2020). Against this backdrop, we contend *all kindergarten students* can learn complex science content and achieve a clear pathway toward success in this critical discipline when supported in engaging and developmentally appropriate ways.

Our project will also address **Competitive Preference Priority 1** (*Promote Partnerships with Entities Underrepresented Under the Education Innovation Research Program*) by leveraging an existing line of research that aims to equip early elementary classrooms with high-quality science programs designed to address educational inequities in science for students from low-income backgrounds and underrepresented communities (Doabler et al., 2017; Doabler et al., 2023; Gersib et al., 2023; Rojo et al., 2023). This is relevant because the proposed research will work with kindergarten students from Pine Bluff, Arkansas. Pine Bluff faces significant inequities. For example, the Arkansas Department of Education (ADOE, n.d.) reports that 71% of families in the Pine Bluff area are in current economic distress, with 24% living below poverty thresholds (Census.gov, n.d.). Moreover, on the third-grade Arkansas state science assessment, over 55% of students from Pine Bluff districts scored below the passing threshold (ADOE, n.d.). Further, there are major staffing issues at early childhood centers in central Arkansas, impacting opportunities to build academic language, number sense skills, and scientific inquiry for future Pine Bluff kindergarten students (McKenzie & McGee, 2022). Recognizing these inequities, we intend to serve more than 1,200 “high-need” students in this community. In this proposal, we operationally define high-need students as kindergarten students from low-income backgrounds who enter school at significant risk for low academic performance and are members of groups traditionally underrepresented in the Education Innovation Research program based on race and ethnicity. We apply this broader definition to support a range of kindergarten students who face educational inequities in science.

The current project will also attend to **Competitive Preference Priority 1** by forming a cadre of multidisciplinary researchers from The University of Texas at Austin (UT), which is designated as a Hispanic-Serving Institution (HSI); the University of Arkansas at Pine Bluff

(UA-PB), a public institution of higher education recognized as a Historically Black College and University (HBCU); the University of Virginia (UVA); and WestEd. Given the scope of work called for in the EIR Early Phase RFA to *increase the volume of projects and partners from HBCUs and minority serving institutions*, we deem participation of UT and UA-PB as significant. Both employ current team members who represent groups that have been underrepresented in the EIR program and STEM careers based on race and ethnicity.

A.1 Need for Equitable Science Instruction in Kindergarten Classrooms. In 2020, the National Science Board (NSB) set a vision for ensuring the U.S. remains globally competitive in STEM fields. At the forefront of the NSB's Vision 2030 was offering equitable and unparalleled STEM learning opportunities for *all* students regardless of socioeconomic or demographic background. Despite Vision 2030's focus on K-12 STEM education, results from the 2019 NAEP Science assessment suggest that too many U.S. students fall short of national benchmarks measuring proficiency of key STEM concepts and practices. For example, 64% of fourth-grade U.S. students in science scored below the 2019 NAEP Proficient achievement level (NCES, 2020). NAEP science data suggest a bleaker picture for fourth-grade students from Arkansas, where over two-thirds of fourth-grade students did not reach basic proficiency levels in science (NCES, 2016).

Indications of inequity in science appear early, with data from the Early Childhood Longitudinal Study – Kindergarten (ECLS-K) suggesting such opportunity gaps surface upon kindergarten entry and persist into the later grades (Morgan et al., 2016; Claessens et al., 2009). Findings from the ECLS-K dataset further indicate that kindergarten students identifying as Black, Hispanic, and students from low-income backgrounds are disproportionally affected by these disparities in science achievement. Here, we briefly address three sources of inequity that can be implicated as contributing factors to the struggles many high-need kindergarten students experience in science. Above all, is the shortage of opportunities to build receptive and expressive language within the domain of science (Doabler et al., 2021). The lack of opportunities to hear and use academic language can hamper high-need students from meaningful participation in science instruction and discourse (Burchinal et al., 2022). Many kindergarten students also receive little support to develop early number sense skills prior to school entry (Duncan et al., 2007; Morgan et al., 2009). Lacking a robust sense of number and operations can result in students struggling to solve science-related problems. Another source of inequity is the depth of background experiences required to connect new information with previously acquired knowledge. Research suggests high-need students receive fewer informal opportunities than their peers to build background knowledge on science (Morgan et al., 2009). Collectively, these sources of inequity spotlight the urgent need to invest in kindergarten science instruction. *High-quality science instruction in kindergarten has the potential to engage high-need students in complex science content and thus offer them problem-solving skills that generalize to the later grades and situations outside of school*.

A.2 Using Innovation to Reduce Opportunity Gaps in Science for High-Need Students. Our project is novel and innovative for several reasons. First, the Sci-K program will adhere to the recommendations of the NGSS (NGSS Lead States, 2013) by promoting a "three-dimensional approach" to science learning (National Research Council [NRC], 2012; Nordine & Lee, 2021). In this way, it will support students in making direct connections among the different dimensions of science (i.e., science and engineering practices, core ideas, and crosscutting concepts) to explore and understand the world around them. This three-dimensional approach vastly differs from current practice in kindergarten classrooms, where science is scarcely taught, and when instruction is delivered, science ideas, concepts, and practices are often taught in isolation (NCES, 2020).

Second, we will design the Sci-K program to offer teachers an instructional tool to meet diverse learning needs. While a substantial body of evidence suggests students' understanding of science content and practices is aided by intentional instructional support (Alfieri et al., 2011; NRC, 2012; Therrien et al., 2017), few science programs incorporate validated principles of instruction, such as scaffolded student practice opportunities (WWC, 2023). Recognizing that the level of instructional guidance required to learn core ideas, concepts, and practices varies by student, Sci-K will offer the versatility to successfully differentiate science instruction based on a student's strengths or learner characteristics. For example, because many students need support to represent and interpret data (Doabler et al., 2021; Litke & Hill, 2020), Sci-K will provide teachers with guidelines to demonstrate how to work with quantitative and qualitative information, such as describing weather patterns over time.

Third, our project differentiates from current science education practice because it will use science instruction as a backdrop to concurrently build students' academic language and early number sense skills. This integrated approach will support students' expressive and receptive language and allow them to use language for authentic purposes—communicating their understanding of science in increasingly precise ways (Grapin et al., 2019). Moreover, it will serve as context for students to apply their early numeracy skills as a demonstration of their understanding of science.

Finally, and above all, evidence-based science programs are in short supply for kindergarten classrooms. For example, of the 2022-2023 state-approved kindergarten science programs from California, Florida, and Texas, three states with considerable sway on the national textbook market, none have empirical research that meets the WWC (2023) screening criteria. To develop Sci-K, we will prioritize three core components that our team validated within a second-grade science program. These components, which are described in detail in Section B, include: **Three-Dimensional Learning Activities**, **Validated Principles of Instruction**, and **Effective Professional Development (PD)**. When intentionally integrated, these components facilitate opportunities for students to: (a) use physical models to discover and explore phenomena, (b) learn and use new academic language during science discourse, and (c) apply mathematics concepts to solve real-world problems.

To build an evidentiary basis for the core components of our second-grade science program, we conducted two methodologically rigorous studies in over 30 second-grade classrooms involving nearly 500 students, many whom received special education services, identified as English learners, and were from low socioeconomic (SES) backgrounds (Doabler et al., 2021; Gersib et al., 2023). Results were promising, with effect sizes (Hedges' *g*) on STEM-related outcomes ranging from 0.48 to 0.94. Relevant to the proposed research, results also indicated that all students benefited from our science program, regardless of their gender, SES status, eligibility for special education, identified race and ethnicity, and initial academic skill levels (Doabler et al., 2021; Rojo et al., 2023). Thus, because our program resulted in equitable outcomes for some of the most vulnerable students, including students with marginalized social identities, we will adapt its three core components to make them developmentally appropriate for kindergarten students and in turn use them as guideposts for designing the Sci-K program.

B. QUALITY OF THE PROJECT DESIGN

B.1 Conceptual Framework. Figure 1 in Appendix G presents our Logic Model, depicting how the implementation of the Sci-K program supports high-need students in meeting the NGSS performance expectations for kindergarten. Our logic model posits that when kindergarten teachers deliver the program with strong implementation fidelity, the program will lead to three-dimensional science learning, increased academic language, and early number sense development. To address these targeted student outcomes, the program and its corresponding PD workshops will support teachers' use of validated principles of instruction and facilitation of high-quality instructional interactions around three-dimensional science learning. These interactions will allow students to work independently and with their peers when building and exploring models about the world. The mechanism of change includes three validated components of science instruction, each described below and depicted in Table 1.

Component 1 –Bundles of Three-Dimensional Science Learning: The first component underlying the Sci-K program is a targeted focus on three-dimensional learning as outlined in the NGSS performance expectations (PEs) for kindergarten (NGSS Lead States, 2013). A robust and lasting understanding of three-dimensional learning is essential for students' development of science proficiency (NRC, 2012). As such, our design team will assemble 18 weeks of science instruction into three "bundles" (NGSS Lead States, 2013) focused on the three-dimensional nature of the NGSS. *Bundle-1 Physical Science* will provide 4 weeks of instruction focused on four PEs (K-PS2-1, K-PS2-2, K-ESS2-1, K-2-ETS1-3); *Bundle-2 Life Science* will offer 8 weeks of instruction on seven PEs (K-LS1-1, K-ESS2-1, K-ESS2-2, K-ESS3-1, K-ESS3-2, K-ESS3-3, K-2-ETS1-1), and *Bundle-3 Earth and Space Science* will provide 6 weeks of instruction on four PEs (K-PS3-1, K-PS3-2, K-ESS2-1, K-2-ETS1-2). We prioritize these PEs to offer opportunities for kindergarten students to explore various science phenomena relevant to later science learning and their everyday lives. Each bundle will target a NGSS-recommended guiding question. Bundle-1 (Physical Science) will address: *How do objects move and what happens when they* *interact?*; Bundle-2 (Life Science) will address: *What is the relationship between the needs of different plants and animals and the places they live?*; and Bundle-3 (Earth and Space Science) will address: *What can we observe about sunlight?*

[redacted]

Component 2 – Validated Principles of Instruction: The second component entails validated principles of instruction (Burchinal et al., 2022; Fuchs et al., 2021; Hughes al., 2017).

Compelling evidence from national assessments suggests that many elementary students require intentionally designed and delivered instruction to acquire three-dimensional learning in science (NCES, 2020). Therefore, the Sci-K program will incorporate validated principles of instruction, such as purposefully selected instructional tasks and activities and teacher-provided academic feedback (Therrien et al., 2017). Additionally, the Sci-K program will include various physical models to explore relevant phenomena and encourage students to communicate their understanding through multiple modalities, including gestures, science talk, and drawings.

Component 3 – Effective Professional Development: Because active participation is integral to the learning process, this final component actively engages teachers in the Sci-K materials, offering opportunities to practice teaching sample activities and receive feedback from key project staff. To ensure high degrees of implementation fidelity, teachers will receive in-class coaching support from curriculum experts with specialized knowledge and training in elementary science instruction and the Sci-K program. Our coaching model will be based on procedures developed and validated in our previous efficacy trials (e.g., Doabler et al., 2016).

[redacted]

[redacted]

[redacted]

2. Validated Principles of Instruction	
2a. Intentionally Planned Activities	Purposefully selected and sequenced activities that facilitate opportunities for students to practice.
2b. Instructional Interactions	Frequent opportunities for all students to explore and reach three-dimensional science learning.
2c. Academic Feedback	Opportunities to reinforce students' learning of science content and activities.
3. Effective Professional Development	
3a. Workshops: Bundles 1-3	Three 6-hour PD workshops on content and pedagogical knowledge for teaching science.
3b. Implementation Support	In-class coaching visits to increase implementation fidelity and high-quality instructional interactions.

B.2 Yearly Objectives and Outcomes. The project's yearly objectives and outcomes are centered around three stages of work: Sci-K Development and Refinement, Feasibility Testing, and Impact Study (see Table 2). As depicted, we will develop Bundles 1-3 of the Sci-K program across the project's first three years, with Year-1 (2024) targeting Physical Science, Year 2 (2025) targeting Life Science, and Year-3 (2026) targeting Earth and Space Science. We will also develop assessment items for each bundle to determine individual student achievement toward the targeted PEs. Following each stage of development will be a Rapid Activity Testing Experiment (RATE) to test the bundles' feasibility, usability, and likeability. We will refine the components of each bundle using formative data (e.g., teacher interviews) collected in the respective RATE. The final years of the project will involve an impact study designed to test the impact of the complete Sci-K program on targeted teacher and student outcomes.

Sci-K Development and Refinement: We contend that critical to the promise of science programs to increase STEM-related outcomes for high-need students is the manner in which they are designed. Therefore, to craft the Sci-K program, we will employ a proven, replicable design methodology grounded in a series of iterative *development, testing* and *revision* cycles (Brown, 1992; Clements, 2007; Cobb et al., 2003; Doabler et al., 2015a). Formative data collected during the cycles will be used to increase the program's quality; obtain estimates of its feasibility, usability, and likeability; and ensure that Bundles 1-3 can be integrated as a complete program in the project's impact study. The effectiveness of this methodology is well documented in our numerous mathematics and science programs funded by the National Science Foundation (NSF), Institute of Education Sciences (IES), and the Office of Special Education Programs (OSEP) (e.g., Doabler et al., 2013; Doabler et al., 2015b; Doabler et al., 2017; Hand, Therrien et al. 2009; Therrien et al., 2010). Notably, the value of our development work has been realized through the procurement of subsequent federal funding to test the impact of many of these programs in large-scale efficacy trials (e.g., Clarke, Doabler, et al., 2016; Doabler et al., 2021; Doabler et al., 2023; Fien, Doabler, et al., 2016; Powell & Doabler, 2020). Taken together, we contend our strong track record in developing and testing STEM programs positions our project team to successfully achieve the objectives and outcomes outlined in this Early Phase project.

We will begin development of Sci-K by adjusting the core components established and tested in our prior work (Doabler et al., 2021) to become developmentally appropriate for kindergarten students, such as embedding more hands-on opportunities and reducing the length of activities to maintain student engagement (Burchinal et al., 2022). The revised components will serve as a foundation for developing each bundle (i.e., Physical Science; Life Science; Earth and Space Science). Because teachers are the fabric of curriculum implementation, we consider them uniquely positioned to provide feedback on the Sci-K program's developmental appropriateness and to troubleshoot existing implementation barriers. Therefore, as in our prior development work, we will empower teachers as partners in the development of Sci-K to enhance community ownership or "buy-in" and help to connect to the experiences students bring to the classroom.

Feasibility Testing: In this project, we will conduct three RATES, the first planned for testing the feasibility of Bundle-1 (Physical Science) in fall/winter of Years 1 and 2, Bundle-2 (Life Science) in fall/winter of Years 2 and 3, and Bundle-3 (Earth and Space Science) in fall/winter of Years 3 and 4. For each RATE, we will recruit a new sample of 10 kindergarten teachers (N = 30) across multiple campuses. Each sample of participating teachers will deliver instruction from the targeted bundle. Two research questions will guide each RATE: (1) *Which features of the bundles appear to maximize student learning and facilitate high-quality instructional interactions?* And (2) *To what extent can teachers feasibly implement the Sci-K bundles?* Each RATE will employ formative observations, surveys, and teacher interviews/focus groups to address the two research questions. When observing Sci-K implementation, we will document the duration of instruction and identify "usability bottlenecks," such as issues with the selected science models. At the end of each RATE, WestEd independent evaluators will conduct teacher interviews and administer surveys to gather formative data to share with our design team for subsequent curricular revisions.

Impact Study: In Years 4 and 5, we will test the impact of the Sci-K program to improve students' three-dimensional science learning, academic language, and early numeracy skills. Specifically, we will conduct an 18-week impact study in 30 kindergarten classrooms from 20 elementary schools in districts located in Pine Bluff, AR. These classrooms, which will be new to the project's sample, will be randomly assigned by WestEd evaluators to treatment (*Sci-K*) or control (business-as-usual) conditions. Details of our impact study are specified in Section E.

Table 2. *Project Objectives and Desired Outcomes*

Objectives	**Outcomes**
Years 1 and 2. Develop and Test Bundle-1	
1.1 Design Bundle-1 (*Physical Science*) instructional and training materials	**1.1** Materials for Bundle-1 fully developed
1.2 Design Bundle-1 Assessment items and FOI Checklist	**1.2** Bundle-1 outcome measures developed
1.3 Conduct RATE #1 (Bundle-1) • 10 classrooms / 200 students	**1.3** Successful RATE #1 conducted in 10 kindergarten classrooms
1.4 Collect formative data in RATE #1 and revise Bundle-1	**1.4** Formative feedback from RATE #1 collected via surveys and observations
Years 2 and 3. Develop and Test Bundle-2	
2.1 Design Bundle-2 (*Life Science*) instructional and training materials	**2.1** Materials for Bundle-2 fully developed
2.2 Design Bundle-2 Assessment items and FOI Checklist	**2.2** Bundle-2 outcome measures developed
2.3 Conduct RATE #2 (Bundle-2) • 10 new classrooms / 200 students	**2.3** Successful RATE #2 conducted in 10 kindergarten classrooms (new sample)
2.4 Collect formative data in RATE #2 and revise Bundle-2	**2.4** Formative feedback from RATE #2 collected via surveys and observations
Years 3 and 4. Develop and Test Bundle-3	
3.1 Design Bundle-3 (*Earth and Space Science*) instructional and training materials	**3.1** Materials for Bundle-3 fully developed
3.2 Design Bundle-3 Assessment items, FOI Checklist, & STEM Vocabulary measure	**3.2** Bundle-3 outcome measures developed
3.3 Conduct RATE #3 (Bundle-3) • 10 new classrooms / 200 students	**3.3** Successful RATE #3 conducted in 10 kindergarten classrooms (new sample)

3.4 Collect formative data in RATE #3 and revise Bundle-3	**3.4** Formative feedback from RATE #3 collected via surveys and observations
3.5 Integrate Bundles 1-3 to form the fully-specified Sci-K program	**3.6** 18 weeks of Sci-K instructional materials focused on kindergarten NGSS PEs
Years 4 and 5. Conduct Impact Study of Sci-K	
4.1 Recruit sample of 30 kindergarten classrooms in Arkansas for impact study	**4.1** District leaders and teachers agree to participate in yearlong impact study
4.2 Randomly assign kindergarten classrooms to treatment or control conditions	**4.2** Participating classrooms in both conditions have baseline equivalence on key teacher and student characteristics
4.3 Provide Sci-K teacher PD workshops	**4.3** Sci-K professional development workshops delivered for each bundle
4.4 Implement Sci-K program	**4.4** Sci-K is implemented with high fidelity & quality in 15 treatment classrooms
4.5 Assess Sci-K impact on all outcomes	**4.5** Data are collected at specified time points and analyzed as planned
4.6 Disseminate project findings	**4.6** Findings are peer reviewed and presented to relevant audiences
4.7 Finalize all Sci-K materials	**4.7** Sci-K teacher, student, and training materials provided to participating districts

B.3 Extent to Which the Design Meets the Needs of the Target Population. The Sci-K program aims to provide equitable opportunities in early science instruction for high-need kindergarten students. To that end, we plan to recruit kindergarten classrooms from school districts in Pine Bluff, AR, where 24% of families live below poverty levels (Census.gov, n.d.). Collectively, we will work with more than 1,200 students across 60 kindergarten classrooms in 20 elementary schools from three high-need school districts (i.e., Pine Bluff, White Hall, and Friendship Aspire Academy; see Appendix C). On the third-grade Arkansas state science assessment, over 55% of students in these districts scored below the passing threshold, indicating significant opportunity gaps in science (ADOE, n.d.). These educational disparities in science disproportionately affect central Arkansas students who are Black, Latino/a, economically disadvantaged, and eligible for special education. Thus, we find it imperative to supply teachers who work with high-need students from central Arkansas with an early science program that addresses these inequities and promotes success in science for all.

To meet this need, we will ground the Sci-K program in validated components (Doabler et al., 2021; Gersib et al., 2023) that prioritize critical science content and the requisite skills in science, academic language skills, and mathematics needed for achieving three-dimensional science learning. Further, the Sci-K program will judiciously integrate validated principles of instruction to best accommodate the strengths and assets of participating students, such as teachers offering students specific academic feedback after science practice opportunities (Burchinal et al., 2022; Fuchs et al., 2021). Our own research supports the impact of these principles in STEM programs, where results from recent efficacy trials indicate significant treatment effects on STEM-related outcomes across genders, ethnicity, SES, and academic skill levels (Doabler et al., 2016; Doabler et al., 2019; Doabler et al., 2023; Rojo et al., 2023).

Further, we will target academic language proficiency and early number sense through integrated STEM learning opportunities. Central Arkansas reported 65% of students did not meet benchmarks in reading and 42% in mathematics by third grade, when state standardized testing begins (ADOE, 2022). Given the impact of early learning opportunities in kindergarten and later schooling (Quinn & Cooc, 2015; Morgan et al., 2016), we will intentionally integrate evidence-based language and mathematics instruction in Sci-K to maximize learning opportunities for high-need kindergarten students.

While equitable science outcomes are our primary focus, we also intend to provide high-quality PD and in-class coaching support to improve teachers' science instructional practices. Specifically, we will train teachers to use validated principles of instruction (e.g., scaffolded verbalization opportunities) to address the strengths and assets of high-need kindergarten students. A major focus of our PD workshops is the active participation of teachers during the learning process (Blank et al., 2008; Garet et al., 2001; Manz & Suarez, 2018). In this way, workshops provide opportunities for teachers to practice facilitating hands-on exploration activities and managing class-wide science conversations.

C. QUALITY OF PROJECT PERSONNEL

Our team comprises a diverse group of curriculum developers, biologists, science educators, early childhood experts, special educators, and former kindergarten teachers. The proposed research will leverage our team's combined knowledge in STEM education and work with high-need students. Collectively, members of our team have successfully engineered and empirically tested more than 10 STEM-related programs for high-need students (e.g., Doabler et al., 2013; Doabler et al., 2015b; Doabler et al., 2017; Hand, Therrien et al. 2009; Therrien et al., 2010). Additionally, our team's extensive experience in designing and implementing interventions for current, former, and non-English learners (ELs) will add significant value to developing the Sci-K program and its academic language component (e.g., Doabler et al., 2016; Martinez et al. ,in press; Vaughn, Martinez, et al., 2009, 2017).

Leadership of the project will be a consortium of nine partners from four institutes. [redacted] **(UT)**, whose research focuses on STEM programs for high-need students, will lead the project. [redacted] has successfully directed or co-directed eight federally-funded design and development projects and seven large-scale efficacy trials involving high-need students. [redacted] will oversee all aspects of the project and lead efforts to integrate number sense instruction into Sci-K. [redacted] **(UA-PB)** will bring expertise in the area of science instruction for students from marginalized groups. [redacted] currently serves as Dean of the College of Education at UA-PB and will lead the outreach efforts with participating districts. [redacted] **(UT)** who specializes in academic language development in the content areas will oversee efforts to design the language features of the Sci-K program. She currently serves as Co-PI of an IES-sponsored R&D center to develop and test science interventions to improve content area learning for ELs. [redacted] **(UVA)** specializes in curriculum development in science and will oversee development of the Sci-K program. [redacted] has successfully directed or co-directed 15 federal grants focused on early science and literacy instruction. [redacted] **(UVA)** is a science educator whose expertise is in science instruction for typically-achieving students. [redacted] will oversee all facets of the exploration activities. [redacted] **(UT)** is a post-doctoral fellow who focuses on evidence-based practices for at-risk learners. [redacted] will oversee efforts to embed validated science teaching and learning techniques in Sci-K. [redacted] **(UA-PB)** is an Associate Professor in UA-PB's Department of Biology. [redacted] will bring expertise in science education for high-need students and knowledge of participating Pine Bluff (AR) schools. [redacted] **(WestEd)** will lead the WestEd evaluation team. He has conducted numerous large-scale evaluation projects and currently serves as Principal Investigator of an EIR Expansive project. [redacted] **(WestEd)** will co-lead the evaluation and oversee implementation fidelity and performance feedback activities.

Our project features an influential group of experts that will serve as a Technical Advisory Panel (TAP) in science, mathematics, literacy, and early childhood education (see Appendix C). The TAP will provide annual feedback on the curriculum development process, with a special emphasis on early childhood education ([redacted] kindergarten science instruction ([redacted] science education for high-need students ([redacted] academic language ([redacted] and early numeracy ([redacted] Further, The Meadows Center for Preventing Educational Risk (MCPER) at UT will support the Sci-K project. MCPER is a leading multidisciplinary research center with a built-in infrastructure that assists in designing curricular materials and coordinates outreach through a repository of up-to-date research findings and practitioner tools. In the past 20 years, MCPER has worked with tens of thousands high-need students in hundreds of U.S. schools.

D. MANAGEMENT PLAN

To monitor the team's progress toward our specified objectives, we created a comprehensive management plan (see Table J.1, Appendix J). By project year, the plan lists the key objectives and their associated milestones and tasks. Participating institutes (i.e., UT, UVA, UA-PB, and WestEd) responsible for each milestone and task are also noted. Central to meeting the specified objectives and milestones will be a communication plan that includes weekly project meetings with participating institutes along with the use of cloud-monitoring tools (e.g., Airtable) that will track our overall progress. As depicted in the management plan, the project's first three years will focus on objectives related to developing Bundles 1-3, PD workshop materials, and project surveys and prototype items for the bundle assessments. Our team will also utilize the first three years to prepare for and conduct the three RATES. Each RATE will be conducted using a new sample of 10 participating kindergarten classrooms. Data collected during the RATES will be used by our team to refine the bundles and integrate them into the complete Sci-K program. A primary objective of the project's fourth year is preparing for the impact study to be conducted across the project's final two years. Impact of the fully-developed Sci-K program on targeted teacher and student outcomes will be tested in a new sample of 30 kindergarten classrooms involving over 600 high-need students. The milestones associated with this objective include training treatment teachers to implement the Sci-K program, collecting student outcome data at specified time periods, conducting direct observations of program implementation and student science learning, administering teacher surveys, and facilitating teacher exit interviews / focus groups. The remainder of the project's final year will entail data analysis, dissemination, and reporting. To promote sustainability of Sci-K in treatment and control classrooms, the project will conclude by supplying participating schools with all Sci-K teaching and PD materials.

E. QUALITY OF PROJECT EVALUATION

WestEd will lead an independent evaluation of Sci-K, including process, implementation, cost, and impact analyses to address evaluation questions that prioritize the Standards for Excellence in Education Research (SEER; https://ies.ed.gov/seer/). WestEd has conducted numerous multisite randomized controlled trials for EIR, IES, NSF, and other federal and state organizations. The Sci-K evaluation principal investigator (PI), [redacted] is currently PI of a 2022 EIR Expansion project evaluation and lead methodologist for a 2022 EIR Mid-Phase project. Given these roles, he has deep knowledge of EIR evaluation approaches.

The evaluation will include studies of (a) the impact of Sci-K on confirmatory outcomes, using a design that meets WWC 5.0 Standards Without Reservations, preregistered in the Registry of Efficacy and Effectiveness Studies (REES) (SEER1); (b) fidelity of implementation (FOI) (SEER3 & SEER4); (c) process studies (i.e., RATE) with rapid-cycle feedback to inform Sci-K about FOI and factors that facilitate or impede program development, scaling, and potential replication (SEER8); and (d) a cost analysis and cost effectiveness study (SEER5) using the ingredients method (Levin et al., 2017) to support sustainability and to understand how resources may be directed to achieve maximum benefit.

E.1 Evaluation Questions. The evaluation will address questions concerning the implementation of key program components and confirmatory and exploratory impacts on intermediate and final outcomes.

Table 3. *Evaluation Questions and Data Sources*

Evaluation Question	Data Sources
Are fidelity of implementation thresholds reached?	Teacher surveys, implementation logs, classroom observations, Sci-K Fidelity Checklist
What are the barriers and supports to successful implementation?	Teacher and developer surveys and interviews
What is the achieved treatment-control contrast?	Teacher surveys regarding Sci-K implementation in both conditions
Confirmatory Impact Question	
Is there a positive intent-to-treat impact of Sci-K, relative to business-as-usual, on kindergarten students': 1. Three-Dimensional Science Learning 2. Science vocabulary 3. Early number sense skills	ECLS-K General Knowledge Test (GKT; https://nces.ed.gov/ecls/) Sci-K Bundles Assessment (to be developed) Vocabulary Assessment (to be developed) Assessing Student Proficiency in Early Number Sense (ASPENS; Clarke et al., 2011)
Exploratory Impact Questions	
Impacts on potential mediators Does implementation fidelity, teacher science practices, science discourse, and instructional interactions mediate the impact of Sci-K on confirmatory outcomes, including Three-Dimensional Science Learning, science vocabulary, and early number sense?	Sci-K FOI Checklist (α = NA) NGSS Science and Engineering Practices Survey (Kang et al., 2018) (α = nr) Scientific Discourse Instrument (SDI; Osborne et al., 2019).
Moderating/differential impacts Is there a differential impact of Sci-K depending on race/ethnicity, gender, disability status & specific disability, EL status, free/reduced-price lunch status, or specific school enrollment?	Student demographic data

Note. Each key data source in Table 3 is described fully in Appendix J (see J.3).

E.2. Impact Evaluation That Meets WWC Standards Without Reservations. The confirmatory and exploratory research questions address key program components, main proximal outcomes, and final impact outcomes from the logic model. The confirmatory research question is as follows: Is there a positive intent-to-treat (ITT) impact of Sci-K relative to business-as-usual (BAU) on students' three-dimensional science learning (NGSS Lead States, 2013), academic language, and early number sense skills.

E.2.1 Samples. The impact study will examine the effects of Sci-K on outcomes for 600 kindergarten students in 30 classrooms in diverse, high-need elementary schools in Arkansas.

E.2.2 Randomization. WestEd will randomly assign the 30 classrooms to either the treatment (Sci-K; 15 classrooms) or control (BAU; 15 classrooms) condition using the *blockTools* (Moore, 2012) package in *R*. Randomization will block by classrooms-level characteristics, which may include the percentage of students in the classroom by race/ethnicity, disability status, EL status, free or reduced price lunch (FRPL), years of teaching experience, or certification to ensure that the teachers and their classrooms are equivalent on key characteristics in each condition at baseline. We anticipate that each classroom will include 20 kindergarten students, resulting in 600 total participants. The cluster-level RCT is designed to meet WWC 5.0 standards without reservations. We will exclude all joiners to the classrooms after randomization, per WWC 5.0. Contamination will be monitored using teacher logs and observations. Given the compressed implementation period (one school year), we do not expect teachers to attrite. To further support participation and mitigate classroom-level attrition, IRB-approved financial incentives will be offered to participating teachers.

E.2.3 Statistical Power. WestEd evaluated the minimum detectable effect size (MDES) for confirmatory impacts on proximal student outcomes (see Table 3) assuming a teacher-level RCT, with 30 teachers, and 570 students remaining. We explored multiple scenarios based on these sample sizes, with several plausible assumptions about variance partitioning. We assumed 80%

power, Type-1 error rate 5%, and specific values of the ICC, $\boldsymbol{R^2}$ and other parameters described in Appendix J. The MDES ranges between .233 and .289, and conservatively, we assumed the latter. This effect size is smaller than effect sizes observed in previous studies of impacts of similar programs on similar outcomes (e.g., Chen & Yan, 2019; Taylor et al., 2019).

E.2.4 Impact Measures. Included outcomes are listed in Table 4 (see full description in Appendix J). Confirmatory and exploratory analyses will rely on the following instruments:

Table 4. *Evaluation Measures*

Unit	Domain	Measure	Timing
Students			
	Three-Dimensional Science Learning	ECLS-K General Knowledge Assessment	Pre and Post
	Three-Dimensional Science Learning	Sci-K Bundles (1-3) Assessment	Pre and Post
	Science Vocabulary	Sci-K Science Vocabulary Assessment	Pre and Post
	Early Number Sense	ASPENS	Pre and Post
Teachers			
	FOI	Sci-K FOI Checklist	3 logs (treatment teachers only)
	Science Instruction	NGSS Science and Engineering Practices Survey	Pre only
	Science Discourse / Instructional Interactions	SDI	3 observations per classroom
	Treatment/control Contrast	Teacher instructional logs	Monthly

E.2.5 Impact Analysis. WestEd will use hierarchical linear models (HLM; Raudenbush & Bryk, 2002) applied to cluster-level RCTs (Bloom, 2005) for estimates of intent-to-treat impact. The standard form of the benchmark impact model (detailed in Appendix J) will include an indicator of treatment status, student-level baseline covariates (e.g., race/ethnicity, gender, disability status, EL status, SES, and pre-treatment measures), teacher covariates, fixed effects for school, and teacher and student random effects. To address missing data, we will use the sequential modeling imputation approach (Grund et al., 2021), which uses Markov chain Monte Carlo (MCMC) methods to estimate the parameters of the imputation models and sample imputations for the missing data from the conditional distributions of the variables (Gelman et al., 2014). For the confirmatory impact analyses, we will follow WWC topic-area review protocols to report all necessary statistics, including obtaining sample sizes at each stage in executing the study design, determining baseline equivalence on demographics and pretests, and calculating standardized mean difference effect sizes.

For exploratory analyses, we will assess differential impacts on confirmatory outcomes for important student and teacher moderators (e.g., race/ethnicity, gender, disability status, EL status, SES, years of teaching experience). Moderation models will include interaction effects at the level of moderation (e.g., student [level 1], classroom [level 2]). We will estimate mediators using a multilevel structural equation modeling (ML-SEM) framework. Specifically, we will examine whether classroom-level factors, such as FOI, science discourse, and instructional interactions, mediate the direct effect of Sci-K on student outcomes. To conduct these analyses, we will use *lme4* (Bates et al., 2015), *Lavaan* (Rosseel, 2012), and other packages in *R*. All statistical code will be preregistered in REES.

E.2.6 Cost Effectiveness. WestEd will conduct a cost analysis based on the Resource Cost Model (Levin & McEwan, 2002) to provide information regarding the cost of implementing Sci-K and whether it is cost effective relative to the BAU condition. Costs will be identified in both the Sci-K and BAU conditions using the "ingredients method" (Levin et al., 2017). Analyses will identify the costs associated with each component of the program, distinguish start-up costs from on-going costs, and convert totals to per-student costs. We will then combine the cost information and effect size estimates to describe the impact of Sci-K on a per-dollar basis following the most up-to-date recommendations for cost analyses (Hollands et al., 2021).

E.3 Fidelity of Implementation Evaluation.

E.3.1 Fidelity of Implementation (FOI). To collect FOI information, the impact study will use an incorporated FOI reporting system. This system includes Specific, Measurable, Attainable, Realistic, Timely (SMART) thresholds for monitoring objective performance measures and integrating feedback. This system will rely on the Sci-K FOI Checklist (teacher log and direct observation versions) to be developed during the project's first three years. WestEd will assess adherence to an on-going adaptation of the program logic model (Appendix G), including key components, outputs related to inputs, and attainment of fidelity thresholds (SEER3 & SEER4). Key components and fidelity thresholds include: (a) Sci-K recruits 30 teachers for the RCT, (b) the Sci-K PD workshops and on-going performance feedback is delivered to 100% of teachers; and (c) teachers deliver 90% of the Sci-K program (i.e., each bundle delivered) as measured by the Sci-K FOI Tool. Findings will be regularly shared with the Sci-K design and implementation team to decide whether key components of the program and fidelity thresholds have been met and to make necessary adjustments.

E.3.2 Variation in Implementation. During the impact study, WestEd will collect monthly teacher practice logs from all Sci-K and BAU teachers regarding their instructional practices and routines and will interview a sample of 10 Sci-K teachers to expand on themes in survey responses and to identify barriers and supports to implementation. This information will provide insights to understand barriers and supports in Sci-K implementation (SEER4). Additionally, a survey will gauge teachers' confidence in teaching for three-dimensional science learning (NGSS Science and Engineering Practices Survey, Kang et al., 2018). The teacher log information will be reported to the Sci-K implementation team to support program implementation and to inform the development of a replicable model and refined logic model.

E.3.3 Treatment-Control Contrast. Science coverage data in the Sci-K and BAU conditions will be collected through teacher logs to evaluate (a) the planned and realized treatment–control contrasts (Weiss et al., 2014) and (b) achieved relative strength of the Sci-K program (Hulleman & Cordray, 2009).

E.4 Potential for Sustainability and Scale-Up. Surveys and focus groups of key participants (including Sci-K developers and teachers) will establish the classroom-level conditions for sustaining Sci-K program components (SEER8). This information will inform program adjustments and support scaling for new contexts.

E.5 Evaluation Performance Feedback. A primary goal of the evaluation is to provide frequent performance feedback to project staff and assessment of progress toward intended outcomes that will allow on-going adaptation and improvement of the Sci-K model and its implementation. The RATE studies and implementation of the impact study will allow WestEd evaluators to monitor progress and serve as a critical and independent thought partner, helping the Sci-K team refine its logic model, confirm fidelity thresholds, develop measures, and establish which program components are implemented successfully or in need of refinement. Working together in the RATES, UT, UVA, UA-PB, and WestEd will identify specific questions that are critical to the continuous improvement of the program. The RATE studies will provide opportunities to evaluate implementation of the individual Sci-K bundles (e.g., Life Science). The long-term goals are to refine the Sci-K logic model and to provide data to support a viable and scalable process that is suited to mid-phase validation, dissemination, and scalability.

References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18. https://doi.org/10.1037/a0021017

Arkansas Department of Education. (n.d.). Home - ADE Data Center. Retrieved July 25, 2023, from https://adedata.arkansas.gov/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01 Blank et al., 2008

Bloom, H. S. (2005). Randomizing Groups to Evaluate Place-Based Programs. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 115–172). Russell Sage Foundation.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*(2), 141-178.

Burchinal, M., Bierman, K., Gonzalez, J., McClelland, M., Nelson, K., Pentimonti, J., Purpura, D., Sachs, J., Sarama, J., Schlesinger-Devlin, E., & Washington, J. (2022). *Preparing Young Children for School*. (WWC 2022009). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from http://whatworks.ed.gov/

Chen, C-H., & Yang, Y-C. (2019). Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. *Educational Research Review*, *26*, 71-81. https://doi.org/10.1016/j.edurev.2018.11.001

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth grade achievement: Evidence from the ECLS-K. *Economics of Education Review*, *28*(4), 415-227.

Clarke, B., Rolfhus, E., Dimino, J., & Gersten, R. M. (2011). *Assessing Student Proficiency of Number Sense (ASPENS)*. Longmont, CO: Cambium Learning Group, Sopris Learning.

Clarke, B., Doabler, C. T., Smolkowski, K., Kurtz-Nelson, E., Baker, S., Fien, H., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness*. *9*, 607-634. https://doi.org/10.1080/19345747.2015.1116034

Clarke, B., Doabler, C. T., Turtura, J., Fien, H. (2021). *Building a flexible and comprehensive approach to supporting student development of whole number understanding (Project Flex).* National Science Foundation - Directorate for Education & Human Resources, Research on Learning in Formal and Informal Settings, Discovery Research K-12 Design & Development.

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula". *Journal for Research in Mathematics Education, 38*, 35–70. https://doi.org/10.2307/30034927

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*, 9-13.

Doabler, C. T., Clarke, B., Fien, H., Baker, S. K., Kosty, D. B., & Cary, M. S. (2015a). The science behind curriculum development and evaluation: Taking a design science approach in the production of a tier 2 mathematics curriculum. *Learning Disability Quarterly, 38*, 97–111. https://doi.org/10.1177/0731948713520555

Doabler, C. T., Clarke, B., Kosty, D. B., Kurtz-Nelson, E., Fien, F., Smolkowski, K., & Baker, S. K. (2016). Testing the efficacy of a tier 2 mathematics intervention: A conceptual replication study. *Exceptional Children, 83*, 92-110. https://doi.org/10.1177/0014402916660084

Doabler, C.T., Clarke, B., Nelson, N.J, & Fien, H. (2015b). *Precision Mathematics: Using interactive gaming technology to build student proficiency in the foundational concepts and problem solving skills of measurement and data analysis.* (National Science Foundation, Directorate for Education & Human Resources, Research on Learning in Formal and Informal Settings, Discovery Research K-12.

Doabler, C. T., Nelson, N.J. & Gause, M. (2013-2015). *Development of a second grade game-based integrated learning system to target whole numbers and operations in base ten and operations and algebraic thinking*. U.S. Department of Education; Institute of Education Sciences, Small Business Innovation Research Program – Phase II.

Doabler, C. T., Therrien, W.J., Longhi, M. A., Roberts, G., Hess, K., Maddox, S. A., Uy, J., Lovette, G. E., Fall, A-M., Kimmel, G., Benson, S., VanUitert, V. J., Wilson, S. E., Powell, S. R., Sampson, V. S., & Toprac, P. (2021). Efficacy of a second-grade science program: Increasing science outcomes for all learners. *Remedial and Special Education, 42,* 140-154. https://doi.org/10.1177/0741932521989

Doabler, C. T., Therrien, W., Powell, S., & Sampson, V. (2017-2021). *Promoting scientific explorers among students with learning disabilities: The design and testing of a grade 2 science program focused on Earth's systems.* National Science Foundation, Directorate for Education & Human Resources, Research on Learning in Formal and Informal Settings, Discovery Research K-12, NSF 15-592.

Doabler, C. T., Rojo, M., Gersib, J. A., Fall, A. M., Longhi, M., Lovette, G., Uy, J., Johnson, K., Ghafghazi, S., Phelps, J., Powell, S., & Therrien, W. (2023). (submitted for publication). How do initial mathematics and reading skills impact student science outcomes? *Journal of Learning Disabilities*.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428.

Fuchs, L.S., Newman-Gonchar, R., Schumacher, R., Dougherty, B., Bucka, N., Karp, K.S., Woodward, J., Clarke, B., Jordan, N. C., Gersten, R., Jayanthi, M., Keating, B., & Morgan, S. (2021). *Assisting Students Struggling with Mathematics: Intervention in the Elementary Grades* (WWC 2021006). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from http://whatworks.ed.gov/

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). *What makes professional development effective? Results from a national sample of teachers. American Educational Research Journal, 38*(4), 915–945.

Geary, D. C. (2018). Growth of symbolic number knowledge accelerates after children understand cardinality. *Cognition*, *177*, 69-78.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D., Vehtari, A., & Rubin, D. B. (2014) *Bayesian data analysis* (3rd ed.) CRC press.

Gersib, J. A., Doabler, C. T., Fall, A. M., Roberts, G., Longhi, M., Brafford, T., Johnson, K., & Therrien, W. T. (2023). (submitted for publication). What works in science? A randomized controlled trial evaluating two forms of science instruction.

Grapin, S. E., Llosa, L., Haas, A., Goggins, M., & Lee, O. (2019). Precision: Toward a meaning-centered view of language use with English learners in the content areas. *Linguists and Education* 50, 71–83.

Grund, S., Lutke, O., & Robitzsch, A. (2021). Multiple imputation of missing data in multilevel models with the R package mdmd: A flexible sequential modeling approach. *Behavior Research Methods*, *53*, 2631-2649. https://doi.org/10.3758/s13428-020-01530-0

Hand, B., Therrien, W., & Shelly, M. *The Efficacy of the Science Writing Heuristic Approach*. (2009-2013). U.S. Department of Education; Institute of Education Sciences NCSER Goal 3 Efficacy Trial.

Hollands, F.M., Pratt-Williams, J., & Shand, R. (2021). *Cost analysis standards & guidelines 1.1*. Cost Analysis in Practice (CAP) Project. https://capproject.org/resources

Hughes, C., Morris, J., Therrien, W. & Benson, S. (2017). Explicit Instruction: Historical and Contemporary Contexts. *Learning Disabilities Research and Practice, 32*(3), 140-148.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*(1), 88-110. https://doi.org/10.1080/19345740802539325

Irwin, V., Wang, K., Tezil, T., Zhang, J,m Filbey, A., Jung, J., Mann, F. B., D., R., & Parker, S. (2023). Report on the Condition of Education 2023. NCES 2023-144. National Center for Education Statistics.

Kang, E. J. S., Donovan, C., & McCarthy, M. J. (2018). Exploring elementary teachers' pedagogical content knowledge and confidence in implementing the NGSS Science and Engineering Practices. *Journal of Science Teacher Education*, *1*, 9-29. https://doi.org/10.1080/1046560X.2017.1415616

Krajcik, J., & Reiser, B. J. (2021). How crosscutting concepts, disciplinary core ideas, and science and engineering practices work together in the classroom. In J. Nordine & O. Lee (Eds.). *Crosscutting concepts: Strengthening science and engineering learning* (pp. 19-42). Arlington, VA: National Science Teaching Association Press.

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2017). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis* (*3rd ed.*). Sage.

Levin, H. M., & McEwan, P. J. (2002). *Cost-effectiveness and educational policy: 2002 yearbook of the American Education Finance Association*. Eye on Education.

Litke, E. & Hill, H. C. (2020). Exploring central tendencies: The teaching of data and statistics in elementary mathematics classrooms. *The Elementary School Journal*, *120*(3), 480-506.

Manz, E., & Suárez, E. (2018). Supporting teachers to negotiate uncertainty for science, students, and teaching. *Science Education*, *102*(4), 771-795.

Martinez, L. R., Fishstrom, S., Vaughn, S., Capin, P., Carlson, C., Andress, T., & Francis, D. (in review). Supporting knowledge and language acquisition for secondary emergent bilinguals through social studies instruction. *Reading Research Quarterly*.

McKenzie, S. C., & McGee, J. B. (2022). *The State of Early Childhood Education & Pre-Kindergarten in Arkansas*. Arkansas Department of Education

Mercer, C. D., & Mercer, A. R. (2005). *Teaching students with learning problems*. Pearson/Merrill/Prentice Hall.

Moore RT, Schnakenberg K (2023). blockTools: Block, Assign, and Diagnose Potential Interference in Randomized Experiments. R package version 0.6.4, https://CRAN.R-project.org/package=blockTools.

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, *42*, 306–321.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher*, 45, 18-35.

National Center for Educational Statistics. (2016). National assessment of educational progress 2015 fourth-grade science assessment. Washington, DC: Institute of Education Sciences.

National Center for Education Statistics. (2020). National assessment of educational progress 2019 fourth-grade science assessment. Washington, DC: Institute of Education Sciences.

National Research Council [NRC]. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Idea.* Washington, D.C.: National Academies press.

National Science Teachers Association. (2014). Early childhood science education. Retrieved from https://www.nsta.org/about/positions/earlychildhood.aspx

Next Generation Science Standards Lead States. 2013. *Next Generation Science Standards: For States, By State*s. Washington, DC: The National Academies Press.

Nordine, J., & Lee, O. (2021). *Crosscutting concepts: Strengthening science and engineering learning*. Arlington, VA: National Science Teaching Association Press.

Osborne, J. F., Borko, H., Fishman, E., Gomez Zaccarelli, F., Berson, E., Busch, K. C., Reigh, E., & Tseng, A. (2019). Impacts of a practice-based professional development program on elementary teachers' facilitation of and student engagement with scientific argumentation. *American Educational Research Journal, 56,* 1067-1112. https://doi.org/10.3102/0002831218812059

Powell, S., & Doabler, C.T. (2020). *The role of algebraic reasoning within additive and multiplicative multi-step problem solving for students with mathematics difficulty*. U.S. Department of Education, Institute of Education Sciences National Center on Special Education Research, Goal 3 Efficacy Grant.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rojo, M., Doabler, C. T., Gersib, J., Fall, A-M., Longhi, M. A., Roberts, G., Kimmel, G. L., Zimmerman, J. U., Lovette, G., Ghafgadzi, S., Powell, S. R., & Therrien, W. J. (2023). Challenging educational inequities in early stem instruction: A case for urgency. *The Elementary School Journal.*

Rosseel, Y. (2012). lavaan: An R package for structural equation nodeling. *Journal of Statistical Software, 48*(2), 1–36. http://www.jstatsoft.org/v48/i02/

Scruggs, T. E., & Mastropieri, M. A. (1993). Special education for the twenty-first century: Integrating learning strategies and thinking skills. *Journal of Learning Disabilities*, *26*(6), 392-398. https://doi.org/10.1177/002221949302600604

Quinn, D. M. & Cooc, N. (2015). Science achievement gaps by gender and race/ethnicity in elementary and middle school: Trends and predictors. *Educational Researcher*, *44*(6), 336-346.

Therrien, W. J., Benson, S. K., Hughes, C. A., & Morris, J. R. (2017). Explicit instruction and next generation science standards aligned classrooms: A fit or a split? *Learning Disabilities Research and Practice*, 32(3), 149–154.

Therrien, W., Hand, B., & Shelly, M. *Science Writing Heuristic Post-Doctoral Fellowship Grant*. (2010). U.S. Department of Education; Institute of Education Sciences.

Vaughn, S., Martinez, L. R., Linan-Thompson, S., Reutebuch, C. K., Carlson, C. D., & Francis, D. J. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness*, *2*(4), 297-324. https://doi.org/10.1080/19345740903167018

Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., & Fall, A. M. (2017). Improving content knowledge and comprehension for English language learners: Findings from a randomized control trial. J*ournal of Educational Psychology, 109*, 22–34. https://doi.org/10.1037/edu0000069

Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management*, *33*(3), 778-808. https://doi.org/10.1002/pam.21760

What Works Clearinghouse (WWC). (2023). *Procedures and standards handbook* (Version 5.0). Institute of Education Sciences, National Center for Education Evaluation and Regional Evaluation, U.S. Department of Education.