Scaling a Project-Based Curriculum that Uses Multiple Literacies to Transform Science Learning in Elementary Schools in the Deep South

Table of Contents

A. Significance

Michigan State University (MSU), in partnership with Alabama A&M University (AAMU), Accelerate Learning Inc. (ALI), and WestEd (WE), is submitting a Mid-phase Education Innovation and Research proposal to the U.S. Department of Education. The proposal addresses **Absolute Priority 1: Moderate Evidence**, **Absolute Priority 3: STEM Education**, and **Competitive Preference Priority 1: Promoting Equity with Partners**. The purpose of this Mid-phase project is to enhance, scale, and evaluate a project-based intervention, *Collaborate Science (ColSci),* designed to promote the teaching and learning of science and literacy and improve motivation and interest in science for high-need students, including historically marginalized groups in STEM in rural and under-resourced elementary schools.

A1. Defined Needs

The challenges and importance of elementary science education. Elementary science education is essential for all children, as it builds foundational science knowledge that is crucial for successful STEM learning in secondary school, future success in college and careers, and for living as productive citizens who can contribute as part of a skilled STEM workforce to keep the U.S. economy competitive (National Academy of Science, Engineering & Mathematics [NASEM], 2022). However, elementary schools face challenges in implementing and sustaining science instruction because many teachers lack the knowledge and pedagogical strategies that align with current research on best practices for science teaching and learning (Trygstad et al., 2020). Teaching science at the elementary level also suffers due to the significant emphasis on language arts and mathematics, which is driven by accountability policies that greatly reduce instructional time for science (Marx & [redacted], 2006; NASEM, 2022). In addition, research shows that motivation and interest in learning science decline by fifth grade, which may contribute to adverse long-term academic outcomes (Bathgate & Schunn, 2016; Ladd & Dinella, 2009).

The challenges of science education in the rural South. There is increasing recognition that early science learning is essential for all students, including those with underrepresented backgrounds in science and science-related fields (e.g., McClure et al., 2017; NASEM, 2022). Students in under-resourced schools often experience far less access and opportunity to learn science in the elementary grades than students from more affluent school districts, thereby creating a discouraging situation of inequitable science access that has potential long-term consequences (Dorph et al., 2011; Marx & [redacted], 2006; Morgan et al., 2016). This problem of elementary science education is exacerbated in underserved urban and rural schools in Alabama. The Alabama State report on science performance at the 4th-grade level for 2023-2024 shows overall dismal performances for all students at the 38th percentile, with Black students substantially lower at the 21st percentile and Whites and Asians scoring substantially higher at the 51st and 61st percentiles, respectively (Alabama State Department of Education, 2024).

A2. Proposed Intervention

We will conduct a mid-phase project to enhance, scale, and perform a randomized controlled trial (RCT) of *ColSci* designed, developed, and tested by researchers from MSU, which ALI distributes. *ColSci* is a full-year digital, print-enhanced Next Generation Science Standards (NGSS Lead States 2013) curriculum that brings science to life by engaging students in making sense of compelling phenomena and finding solutions to practical, real-world problems through a multiple literacies approach grounded in project-based learning (PBL) ([redacted] & [redacted], 2022). The term "multiple literacies" refers to the various disciplinary tools, language, and ways of reasoning that students bring forward as they engage in learning science.

Advantageously, the curriculum integrates science learning with students' literacy development, emphasizing reading, writing, drawing, and oral communication. It uses various texts and media to support knowledge-building and scientific investigation (high-quality trade books, grade-appropriate scientific papers, graphic representations, videos, websites, simulations, etc.). The *ColSci* strategy of integrated science and literacy has significantly improved student outcomes ([redacted] et al., 2022). It has the added benefit of enabling teachers to spend considerably more time on science instruction because it simultaneously enhances learning in both content areas. (See Appendix J6.4 for the literacy connection in the content).

To support science and literacy together, *ColSci* immerses students in meaningful, experiential, project-based learning where they investigate compelling phenomena that relate to students' everyday interests, curiosities, and surrounding communities. This approach to science instruction sparks students' motivation and engagement for learning and deepens their proficiencies with scientific practices and concepts as they apply them to relevant real-world situations. The curriculum materials include educative features that enable teachers to learn pedagogical strategies and develop their science content knowledge (Davis & [redacted], 2005).

For this mid-phase project, fourth-grade teachers in Alabama's elementary schools will receive *ColSci* teaching and learning materials that align with Alabama state science standards. Additionally, teachers will receive sustained professional development to effectively implement evidence-based, integrated science and literacy PBL pedagogical strategies and content. Our intervention will increase a wide range of students' interest in science, opportunities to learn, and achievement in science and literacy as measured by end-of-year state assessments.

A2.1 Absolute Priority 1: Moderate Evidence. Research on PBL has demonstrated a positive relationship between students' science knowledge and standardized test performance ([redacted] et al., 2022). To test the efficacy of *ColSci* – previously referred to as ML-PBL during its initial development ([redacted] & Schneider, 2021) – researchers at MSU conducted an RCT with 2,371 elementary students in various regions of Michigan to determine if the approach improved third-grade students' science learning and aspects of their social and emotional development. The treatment condition significantly outperformed the control students on a test designed from the State of Michigan's standardized science test items, resulting in a 0.277 effect size – a significant treatment effect in education (Kraft, 2020). This positive effect held regardless of racial and ethnic makeup, household income status, or reading level. Study results provide evidence that *ColSci* meets the requirements of the What Works Clearinghouse (WWC) Standards without reservations (see EIR Proposal Evidence Form), establishing the intervention as having moderate evidence meeting absolute priority 1.

Secondly, strong evidence supports *ColSci* required for **Recommendation 2 (**"Integrate oral and written English language instruction into content-area teaching"**)** from the *WWC Practice Guide: Teaching Academic Content and Literacy to English Learners in Elementary and Middle School* (NCEE 2014-4012; Baker et al., 2014). The recommendation was backed up by five studies meeting WWC standards and indicated that integrating oral and written English language instruction into science enhances science knowledge (p.90-91). These studies were conducted on students in pre-K-2 and 5-6 grades. These studies are close to the 4th-grade population proposed for the project. As stated in the practice guide, "the panel believes results from the five studies apply to students in Grades K–8." (p.88). See the *EIR Proposal Evidence Form* for summaries of the studies discussed above and Appendix J1. for additional evidence from other studies.

This project will build additional evidence of the effectiveness of integrating science and literacy within PBL environments to promote usable knowledge, contributing to this body of research by evaluating the effectiveness of *ColSci* in 4th grade. In doing so, this project will tackle several of the educational problems and issues surrounding the implementation of PBL, provide additional and robust evidence for using integrated science and literacy PBL in rural areas, and address how to best support teachers in implementing this instructional approach.

A2.2 Absolute Priority 3: Promoting Equity in Student Access to Educational Resources and Opportunities: STEM. *ColSci* addresses this priority by providing multiple learning opportunities to improve students' science academic performance and interest in science learning, thereby reducing the achievement gap and improving equity.

Integration of Science and Literacy. ColSci promotes equity in access by providing our target students with more opportunities to learn science via integrating science and literacy, which creates greater instructional time for science. Research conducted with upper elementary students suggests that science and literacy integration increase learning in both domains (Lee et al., 2008; Palincsar & Magnusson, 2001; Romance & Vitale, 2001). Moreover, the NGSS supports the integration of science and literacy. The NGSS emphasizes that all students should engage in science learning that reflects science practices, including defending or refuting claims based on evidence, constructing explanations, and communicating information (NASEM, 2022; NRC, 2014). As students investigate to understand and explain phenomena, they learn scientific methods of speaking, reading, listening, writing, and viewing and effectively communicating findings and providing explanations (Lee, Grapin, & Haas, 2018; Pearson, Moje, & Greenleaf, 2010).

Locally Adaptable Learning Contexts. ColSci equips teachers to implement phenomena-driven PBL, enhancing engagement with meaningful and challenging science activities within their local environments. The intervention provides locally adaptable learning contexts, such as investigating how local land formations have formed over the years, to engage a diverse range of students in science learning regardless of their location. This feature allows teachers to situate *ColSci* in their local context to improve the cultural relevancy of the curriculum to their students. This feature helps to increase a sense of relatedness that enhances students' motivation and interest toward learning science.

Access and Equity Design Features. ColSci includes a digital platform for students and teachers and physical materials for hands-on experiences. Students interact with physical materials within a digital workspace; teachers have all necessary resources, such as lesson plans and unit assessments, housed within their digital platform. Designed in HTML5, *ColSci* is (1) accessible on any device with an internet connection and (2) includes technology-embedded features that enable all students to learn regardless of their learning needs or backgrounds. To enhance accessibility, *ColSci* offers features like text-to-speech with adjustable speaking rates and word highlighting, options to control font size and background color for readability, and multiple means for accessing content through texts, videos, pictures, and simulations. An embedded dictionary supports striving readers, and the digital platform allows teachers to seamlessly personalize instruction by assigning activities individually to students. Assessment materials are provided in downloadable, editable formats, enabling teachers to customize questions and answer choices.

ColSci includes robust support for English learners (ELs) by integrating oral and written English language instruction into science teaching. Each lesson guides teachers on leveraging language-rich contextual opportunities to promote student access, participation, language development outcomes, and meaningful scientific inquiry and discourse engagement. All student materials are available in English and Spanish to support diverse learner needs. Students access these equity features through the digital platform.

A2.3 Competitive Preference Priority: Implementers and Partners. Recognizing the critical role and expertise of historically black colleges and universities in advancing science education, MSU and ALI partnered with AAMU and school districts throughout Alabama to implement the *ColSci* intervention with a new and high-need population. The partnership will ensure that enhancements made to the curricula are locally enacted and useable for students and teachers in elementary schools, mainly rural and urban students with low socioeconomic status. This collaboration will provide valuable experiences and knowledge that enhance the learning of the entire group as we work together to improve the materials. This partnership will enable AAMU faculty to expand their research and development capacity to design PBL materials, provide professional learning (PL) to support teachers and conduct large-scale RCT using advanced quantitative methods. For example, two graduate students at AAMU will work with the WestEd team to collect and analyze data during the evaluation study. They will also get experience authoring research manuscripts and presenting at conferences.

B. Strategies to Scale

B1. Strategies for Overcoming Barriers to Scale

B1.1 Strategies to promote pedagogical knowledge of teachers to increase and sustain students' engagement in learning experiences. In the initial development of *ColSci*, we developed specialized supports, called *prompts*, intended to help teachers effectively attend to students' motivation, engagement, and thinking during science instruction. However, we found the prompts underspecified for teachers' use and thus insufficient to support and sustain the interest and motivation of the full range of students in today's classrooms. Our strategy focuses on revising and creating new prompts to overcome this challenge and ensure that all students are consistently supported. We will embed these modified and new prompts into the teachers' guide materials to equip teachers to enhance student motivation and interest. We will provide a range of prompts, including questions, repeated teaching patterns or routines, and strategies that promote productive thinking, which teachers can select and tailor to the students in their class. For example, a teacher could ask, *"How does what we are exploring today relate to what we discussed yesterday?"* The design of these prompts will help teachers support specific discourse strategies and adopt new forms of discourse to provoke students' motivation and interest in learning science. For example, we will design prompts to guide teachers in supporting students in building positive value toward science, improving their confidence in learning science, enhancing their relatedness to science, and building their science identity. For instance, one value prompt a teacher could ask students is: "*How does what they are learning relate to their lives*?" By strategically using prompts, teachers can actively engage students' ideas about, and orientations to, what and how they are learning ([redacted] et al., 2023).

To ensure the effective use of the prompts by teachers, we will enhance the educative features to include statements for teachers that explain why and how to use prompts to strengthen the motivation and interest of a wide range of students, including high-needs students. Moreover, we will conduct PL to support teachers using the prompts to enhance student motivation and engagement. To prepare for this, we will modify the PL materials to include these new features.

B1.2 Strategies to support time and knowledge to use embedded assessment for advancing student learning in science. Evidence shows that providing explicit and actionable feedback tied to the learning goals of embedded assessment promotes student learning ([redacted] et al., 2024). *ColSci* incorporates embedded assessments throughout the curriculum. Unfortunately, previous research indicates that teachers do not use the embedded assessments regularly enough and do not provide feedback effectively to students due to limited time, experience, and/or pedagogical knowledge. To overcome this barrier, we will enact two strategies.

Strategy 1: Support pedagogical knowledge for evaluating student responses. We will design rubrics to support teachers in assessing student responses on embedded assessments. These rubrics will facilitate actionable feedback teachers can provide based on predictable student responses to meet a wide range of student needs. Educative features will include rationales for the rubrics, examples of student responses, example evaluation of student responses, and reasons for the evaluation with suggested actionable feedback to promote teacher learning. We will integrate educative features into the teacher materials to support effectively utilizing the rubrics to understand student responses on the embedded assessments and recommend actionable feedback to promote learning. Additionally, we will provide PL to support teachers using the rubrics to assess student responses on embedded assessments and provide actional feedback.

Strategy 2: Reduce the burden of assessment by automating the feedback loop. ColSci uses a digital platform to collect student responses to the embedded assessments automatically. We will develop a teacher dashboard supported by a backend program to analyze students' responses to embedded assessment tasks on this platform. We will also provide teachers with immediate performance data by individual, class, and aggregate levels. The teacher dashboard will recommend explicit and actionable feedback for teachers to adjust their instruction in response to student needs. This will reduce the time burden for teachers to organize and analyze student responses, shorten the time for giving high-quality feedback to students, and better enable teachers to make "just in time" adjustments to their teaching based on student performance. MSU team members have experience developing machine learning models to analyze students' responses to NGSS-aligned assessment tasks and designing teacher strategies to support informed instructional decisions (He et al., in press). Notably, we will develop machine algorithms using existing secondary student data and open-source generative models (e.g., LLMs from Facebook [Llama 2] (Touvron et al., 2023) and evaluate their basic capabilities ([redacted] et al., 2024). The dashboard and recommended feedback for instructional actions will be co-designed with elementary teachers from our partner districts, and the models will be revised and refined continuously based on feedback from teachers, students, and experts in science education until the final satisfactory outcome is achieved and approved by the teachers.

B2. Management Plan

MSU will lead the proposed five-year project, collaborating with partners AAMU, ALI, and WE. The MSU team will be supported by the CREATE for STEM Institute, which has a track record for developing and testing science education materials and conducting research initiatives that achieve impressive academic outcomes for high-needs students who are culturally and linguistically diverse and from urban and rural school settings. CREATE provides ongoing support for teachers in numerous school districts, and CREATE's researchers have working relationships with district leaders and teachers across the country, including the Deep South.

Cross-institutional members of the project team have worked successfully together on prior federally funded projects focused on the design-based research and evaluation of educational innovations. Team members have expertise in each of the major areas of the project, spanning from design-based studies, research, and development for K-8 classrooms, supporting the implementation of new curricula in schools, scaling of education innovations, and conducting rigorous evaluations, as described below. MSU will manage the overarching project and timeline and lead the development activities with ALI and AAMU to ensure the curriculum will scale. ALI and AAMU will collaborate to recruit schools and support the implementation of *ColSci*. WestEd, with many years of experience in conducting evaluations of educational innovations on multiple Institute of Education Sciences grants, will conduct the independent evaluation of *ColSci* and oversee all aspects of the evaluation study. A technical advisory group with expertise in elementary science education, integrated science and language learning, science education policy, research and practice across southern states, evaluation methodology and WWC standards, and natural language processing will support the team. We have invited experts who agreed to participate. Appendix J2 provides an organizational chart, which identifies the project leadership and illustrates the overarching responsibilities of each organization; Appendix J3 provides details on the capabilities of all key personnel, advisors, and their project roles. In years 1-2, the project team will enhance the existing *ColSci* materials, develop strategies for scaling, and prepare for the evaluation study. In years 3-4, the team will conduct the impact study of *ColSci* in fourth-grade classrooms across 80 elementary schools throughout Alabama (see Appendix C for the supporting letters of school districts). WestEd will lead the impact and implementation fidelity analyses and conduct a cost analysis. In year 5, the team will analyze data, write manuscripts, and disseminate findings (see B4). Appendix J4 provides a detailed timeline, with objectives and responsibilities designated for each year of the five-year project.

B3. Capacity to Bring *ColSci* to Scale

[redacted]**, PI,** is an MSU College of Education Associate Professor. Her research focuses on designing and testing personalized learning environments to promote illstructured problem-solving for a wide range of students, and she has expertise in complex, large dataset research and evaluation studies. [redacted] will oversee all aspects of the project, including acting as the point of contact to the U.S. DOE, overseeing the day-to-day work, ensuring tasks are completed on time and within budget, and communicating and coordinating across all partners (see Appendix J2). [redacted]**, Co-PI,** is a University Distinguished Professor of Science Education and Director of the CREATE for STEM Institute at MSU. He is an internationally recognized expert in science education, science curriculum development, and project-based learning. His research focuses on developing and testing interventions and formative assessments designed to improve the science learning of all students. He will lead the development work to achieve the scaling strategies, oversee the materials' adaptations in collaboration with ALI, and play a key role in dissemination and outreach. [redacted]**., Co-PI,** is the Red Cedar Distinguished Professor of Educational Psychology and Educational Technology and Chairperson of the Department of Counseling, Educational Psychology, and Special Education at MSU. His research focuses on supporting students' motivation and engagement and K-12 schools in implementing digital technologies. He will lead the development work on *ColSci's* motivation features and the assessment instrument for students' motivation and interest in learning science.

[redacted]**, Co-PI,** is Research Manager at Accelerate Learning Inc. She brings expertise in educational psychology and neuroscience related to young learners and in STEM curriculum implementation in elementary schools. She will ensure delivery of all curriculum materials to schools, contribute to outreach to districts for study participation, and co-lead the *ColSci* curriculum implementation in all participating schools. [redacted]**, Co-PI,** Senior Vice President, Research and Measurement at ALI, is a former elementary teacher, school principal, and the founder of the STEMscopes K-8 program. He will work with [redacted] to support outreach and lead the *ColSci* curriculum implementation and the professional learning workshops to prepare teachers for implementing the curriculum.

[redacted]**, Co-PI**, is an Associate Professor of Elementary Education and Coordinator of the PhD Curriculum and Instruction program in the Department of Teacher Education and Leadership at Alabama A&M University. [redacted] will support district outreach and school participation in the study and will be responsible for supporting teachers in the data collection activities. She will oversee AAMU graduate students who will collaborate with WestEd to collect and analyze data as a component of the evaluation study and with MSU to design the rubrics and prompts.

[redacted]**, Co-PI,** is the Research Director of Learning and Technology at WestEd, with more than 15 years of experience in research design and methodology for evaluating the efficacy of educational interventions. [redacted] will lead the program evaluation study's research design, methodology, and analysis. Her research focuses on the design and experimental research of STEM programs and technologies. She is a WWC-certified reviewer and has led several complex, multi-institutional projects as PI or Co-PI. [redacted]**, Co-PI,** is a Senior Director of Science and Engineering Education Research at WestEd. He brings expertise in elementary science education, evaluation of science and engineering programs, design and study of K–12 science assessments and curricula, and classroom-based research on science teaching and learning. He will be responsible for overseeing the evaluation with [redacted]. They will co-direct all components, including study design, random assignment, data collection, impact analysis, reporting, and dissemination.

B4. Mechanisms to Disseminate

We expect that findings from this project will interest and benefit school districts, educators, and educational researchers. The project's leadership team has a rich history of disseminating findings through various communication channels and is well-known for informing multiple stakeholders – teachers, district and state leaders, policymakers, and researchers. To reach our audiences, we will (1) develop a project website hosted by CREATE for STEM at MSU (https://create4stem.msu.edu) with links to AAMU and ALI partners' websites, (2) share our findings at educational conferences (e.g., keynotes and practitioner-focused workshops) and publish papers and book chapters on our research, and (3) disseminate findings with audiences through more contemporary social channels to build broad awareness.

We plan to submit papers to conferences that attract a broad and diverse audience of researchers (e.g., SREE, AERA, NARST, ISLS) and develop presentations for educator conferences like the National Science Teachers Association to disseminate information about outcomes related to curriculum implementation, teacher enactment, and learning outcomes. We will also disseminate findings to our audiences through nationally focused digital channels, including social media, blogs, and e-newsletters. For example, MSU and its partners can leverage their media channels to reach school, district, and state science leaders looking to learn about elementary science curriculum implementation and its impact on learning. We can also utilize professional listservs hosted by the National Science Education Leadership Association (NSELA) and other science practitioner groups, professional learning communities such as the Council of the Great City Schools District Science Leaders, and e-newsletters such as the WestEd NGSS Now E-newsletter that reaches over 29,000 education subscribers.

B5. Utility of Products

ColSci's potential to scale nationally and increase teachers' capability to use the materials effectively directly results from the product development from the latest science teaching and learning research. To ensure utility, the curriculum aligns with the principles of the *Framework for K-12 Science Education* (NRC, 2012) and *Science and Engineering in Preschool Through Elementary Grades* (NASEM, 2022) and the performance expectations of the *Next Generation of Science Standards* (NGSS Lead States, 2013) to improve opportunities for students to learn science and foster interest among high-need students. To facilitate the utility of products, *ColSci* includes (1) *Standards-alignment charts* that illustrate how NGSS performance expectations and ELA Common Core State Standards are covered within and across lessons and over the school year, (2) *Embedded literacy and science connections* throughout the lesson plans with targeted strategies provided to teachers to enable them to support science and literacy together, (3) *Educative features* for teachers to support learning new evidence-based teaching strategies and science content so that teachers from a variety of backgrounds, experiences, and levels of expertise can effectively implement *ColSci* in their classrooms, and (4) *Digital platform* that enables students to have access to high-quality materials that provide learning opportunities to learn challenging science ideas and practices.

C. Quality of the Project Design

C1. Conceptual Framework: Project-based Learning (PBL) and *ColSci*

C1.1 Project-based Learning: PBL is an instructional approach that emphasizes students designing, developing, and constructing solutions to meaningful problems or explaining compelling phenomena ([redacted] et al., 2022). In PBL, driving questions guide students' exploration and sustain engagement as students develop an artifact that demonstrates what they have learned and makes an authentic connection to their communities (see Appendix J5 for PBL features). Consistent with lessening inequality and increasing educational opportunities for all children, PBL reshapes science education by engaging learners in meaningful and robust knowledge-building experiences. Because of its emphasis on exploring relevant questions, PBL promotes collaboration and knowledge-building. Consequently, PBL encourages students to develop the understanding needed for the long-term retention and application of what they have learned (Strobel & van Barneveld, 2009).

C1.2 *ColSci:* The logic model in Appendix G shows the theoretical foundations, each of the key components of *ColSci,* and how these components impact teachers and students. *ColSci* equips elementary teachers to enact project-based learning to emphasize active, rigorous, and content-rich learning aligned to state learning goals. *ColSci* consists of four units covering life science, physical science, Earth and space science, and engineering at each grade level. Because the intervention focuses on students' interests, it is sensitive to the varied needs of their diverse characteristics, including culture, race/ethnicity, and gender. It is also structured to support social and emotional learning, defined in terms of self-reflection, capacity for collaboration, and taking ownership and responsibility for one's work (Jagers et al., 2019).

The units are organized around driving questions with sub-questions that guide day-to-day activities and help students build knowledge. Importantly, *ColSci* is structured to engage students in activities like how *scientists* conduct their work. This includes purposeful attention to collaboration, where students explain, debate, refine, and build knowledge as they work together to address the driving question. Throughout each unit, students are building toward meeting a bundle of NGSS Performance Expectations. Each unit culminates in a final project artifact. See Appendix J6.1–3 for a summary of the performance expectations, fourth-grade learning sequence, and the unit summaries for *ColSci*. A description of features in Unit 1 for fourth grade

– including literacy connections, a table of contents, an example of a student handout, and an example of teacher materials – is shown in Appendix J6.4 – 6.

ColSci provides formative and summative assessments, and the final artifacts include scoring guidance and post-unit assessments. These Post-Unit Assessments use a familiar yet novel phenomenon and offer students an opportunity to apply what they have learned during the unit. An assessment bank is also included where teachers can create customized, standards-based assessments for students to complete.

C2. Clearly Specified and Measurable Goals, Objectives, and Outcomes

We propose the following objectives, strategies, and outcomes/measures (See Exhibit 1).

Exhibit 1. Research Objectives, Strategies, and Outcomes/Measures

Strategies	Outcomes/Measures
Objectives 1. Develop and test strategies to scale and sustain *ColSci*	
1.1 Develop and test rubrics, feedback, and motivational prompts	1.1 Completed rubrics, feedback, and motivational prompts through iteratively design study
1.2 Apply, test, and enhance AI algorithms to code existing student responses; collect new data from 600 students	1.2 Completed AI scoring models of student responses with ~90% reliability in scoring
1.3 Reconfigure a Teacher Dashboard to provide timely student performance data	1.3 Completed fully operational dashboard
1.4 Adapt the teacher PL module for the evaluation	1.4 Completed PL module

1.5 Recruit ten experts and six teachers to provide feedback on rubrics, feedback and motivational prompts, AI scoring, a dashboard, and a PL program 1.6 Pilot enhanced *ColSci* to 6 teachers with their students during four weeks of instruction	1.5 All participants provide feedback during the design studies, report high engagement and ease of use (Survey & Interviews) 1.6 Pilot teachers and students report 100% satisfaction with the feasibility of the enhanced *ColSci* (Surveys & Interviews)
Objective 2. Conduct a rigorous evaluation of *ColSci* to determine impact	
2.1 Prepare research instruments for data collection 2.2 Recruit and randomly assign schools to treatment and control conditions 2.3 WestEd conducts a study of student outcomes, implementation fidelity, and teacher practice 2.4 Complete data analyses and Document findings	2.1 Completed measures and data collection procedures 2.2 80 schools are randomly assigned 2.3 High participation rate by schools w/low attrition; teachers complete all responsibilities 2.4 Fidelity and quality of implementation are above 90%; increase in science performance and literacy growth for treatment students (learning outcome measures), and research questions answered
Objective 3. Disseminate findings and plan scaling and sustainment	
3.1 Create a project website to showcase progress and findings	3.1 The public can access the website

3.2 Reports and manuscripts for researchers, practitioners, and policymakers	3.2 Completed reports; manuscripts submitted to peer-reviewed journals; conference proposals are submitted
3.3 Use of social media and presentations to disseminate nationally	3.3 Share findings through social media and workshop presentations

C3. Addressing the Needs of the Target Population

ColSci aims to bridge the achievement gap by focusing on questions relevant to students' lives. This feature of *ColSci* enhances emotional and cognitive engagement, particularly for underserved students, as students use scientific ideas and practices that foster a deeper understanding and appreciation of science. Students who use *ColSci* will have learning opportunities that result in building a solid foundation of science that is critical to developing scientific understanding to become productive citizens and enter the STEM workforce. Teachers will benefit from science education grounded in cutting-edge research and delivered through rich, meaningful learning opportunities. This project will equip elementary teachers with resources and professional learning to effectively implement PBL in their classrooms. *ColSci* provides comprehensive educational features and professional learning opportunities to enhance teachers' understanding and application of PBL that integrates science and literacy. As a result, teachers acquire the skills and knowledge necessary to create dynamic PBL environments that integrate evidence-based teaching strategies, thereby fostering student achievement in science and literacy. This transformation turns classrooms into active learning spaces, addressing the national imperative to cultivate informed citizens and a skilled, diverse STEM workforce.

In the project's first year (2025), we will conduct a series of design studies to ensure that the teaching prompts, assessment rubrics, dashboard, and machine learning models support student learning. Each design study will implement one curriculum unit (completed in six to seven weeks). We will collect data from 10 science education experts and six elementary teachers in our target Alabama schools. Science education experts and teachers will judge whether the teaching prompts, assessment rubrics, and dashboard align with the learning goal, the accompanying design features, and student and teacher backgrounds. We will ask them to attend to their racial/ethnic, educational, and cultural biases of them. Based on the results, we will revise the features (i.e., prompts, rubrics, and dashboard) for a pilot study.

The pilot study (Spring, 2026) will focus on testing the feasibility of the added features. It will address the following goals: (a) determine the extent to which teachers easily and effectively use the added features to support their science teaching and (b) examine students' responses and engagement to the supporting strategies. Six Alabama teachers will implement one curricular unit over six weeks. The teachers will reflect on how well the added features supported their classroom implementation and will provide suggestions for improvement of the *ColSci.*

D. Project Evaluation

D1. Methods Designed to Meet WWC Standards Without Reservations

WestEd will conduct a rigorous impact study in the form of a clustered randomized controlled trial to address research questions (Exhibit 2) that are aligned with the *ColSci* logic model (see Appendix G). The **independent** evaluation will examine the program's impact on 4th-grade students' science and literacy achievement as well as their interest in science (RQ1-3), the differential effects of *ColSci* on different school contexts or teacher backgrounds, or students' prior achievement or demographics (RQ4), its impact on teachers' science practices and to what extent teacher practices mediate student outcomes (RQ5-6), implementation fidelity and quality

(RQ7-9), and the potential maturation effects of teacher experience with *ColSci* and the replicability of the program (RQ10-11).

Exhibit 2. Summative Research Questions and Data Sources

Research Questions	Data Sources
Impact Analyses	
1. What is the impact of *ColSci* on the science achievement of 4th-grade students?	State standardized science and NGSS-aligned 3-D assessment
2. What is the impact of *ColSci* on the literacy performance of 4th-grade students?	State-standardized ELA and ACCESS EL assessments
3. What is the impact of *ColSci* on student's motivation and interest in science learning?	Study-administered student motivation and interest survey
Moderating and Mediating Analyses	
4. What factors moderate the effects of *ColSci*?	Data for RQ1-3 and relevant background/demographic data
5. What impact does *ColSci* have on teachers' practices?	Study-administered teacher log
6. To what extent do teacher practices mediate the effect of treatment on student outcomes?	Data for RQ1-3 and RQ5
Implementation Fidelity	
7. To what extent are the key components of *ColSci* implemented with fidelity?	PL attendance records, teacher logs, and classroom observation data

8. What are the factors that hinder or facilitate implementation?	Observations, teacher interviews, and ALI facilitator interview data
9. What is the relationship between implementation fidelity and student learning outcomes?	Data for RQ7 and intervention group data for RQ1-3
Exploratory Analyses (maturation effects and replicability)	
10. Is the impact of *ColSci* on student outcomes stronger in teachers' second year of implementation?	Data for RQ1-3
11. When control schools receive the delayed treatment, can teachers implement the program effectively?	Similar data as for RQ7-8

Sample and power. The team will recruit 80 public elementary schools (with at least two teachers in 4th grade), with over 50% classified as rural. We will identify possible stratification variables (e.g., % of students who achieved proficiency, % of minority, % of students in low socioeconomic status) to improve precision and sample balance across groups and develop a plan for recruiting a representative sample to increase the generalizability of the results, with a focus on involving high-need schools. This design will allow participation from approximately 160 teachers and 3,200 students. A power analysis was conducted using *PowerUp!* (Dong & Maynard, 2013) for random assignment at the school level. Assuming an average of two teachers, 15 students per teacher, would remain in the analytic sample at each school, the proposed study has sufficient power to detect effects of 0.168-0.199 standard deviations on student outcomes and 0.381-0.477 for teacher instructional practice (see Appendix. J10 Power Analysis of the Impact Study and J11 Data Analysis Plan and Statistical Model).

Study design. The study will use a school-level, clustered-randomized, controlled experimental design to meet What Works Clearinghouse standards without reservations (WWC, 2022). The RQs will be addressed with data collected from 80 elementary schools during SY 2026-27 (Implementation Year 1) and SY 2027-28 (Implementation Year 2) (see Appendix. J8 for the timeline of evaluation activities). In spring 2026, enrolled schools will be assigned to an intervention group or a comparison group using a stratified random assignment that will consider demographic and prior math and English performance blocked by districts. The 40 schools in the intervention will implement the *ColSci* program in 4th grade over two consecutive school years, where 4th-grade teachers will have two years of experience implementing *ColSci*, each year teaching a new cohort of 4th-grade students. In the comparison group, 4th-grade teachers will continue their current practices (business-as-usual) in both years, using the current district-adopted science materials. They will receive no training with or access to *ColSci*. To incentivize the comparison teachers to engage in the study, the team will offer face-to-face PL focusing on science content, NGSS standards, and science learning in general. After the RCT study, control schools will also have access to *ColSci* in year five.

After randomization, districts and schools will receive assignments, and ALI will begin enrolling intervention schools in its PL program. All 4th-grade teachers who consent to use *ColSci* in the coming year will be included. WestEd will partner with district and school leaders to inform and onboard teachers to the study. ALI will designate one Professional Learning Lead for coordination, oversight, and support and three regional facilitators to offer smaller, more local training. Teachers will receive three days of PL on *ColSci,* PBL, and NGSS before the school year starts. During the school year, 1-day PL sessions will be implemented before each new unit and supplemented with videoconferencing support. Teachers will receive one day of PL refresher during the second implementation year. Regional facilitators will provide continuous support during the school year (see Appendix J7 for *ColSci* PL description).

School-level random assignment is ideal since schools typically implement *ColSci* as a school-wide whole-grade science program. A teacher-level assignment will likely raise the threat of contamination, as teachers in a school tend to discuss and share instructional materials and strategies. The school-level assignment also helps mediate attrition and issues related to teacher turnover and transition. Drawn from prior impact studies by WestEd with school-level randomizations (e.g., Davenport et al., 2019; [redacted] et al., 2023), with the support from ALI and AAMU, we expect low cluster-level attrition, comparable teacher turnover between years, and manageable individual attrition over the two-year study. As incentives, comparison schools will receive the program at no cost at the end of the study. Teachers will receive stipends for PL and data collection activities. WestEd will designate a district coordinator to regularly communicate with district leaders and three teacher liaisons who will serve as points of contact to maintain the engagement level of teachers. A secure database of all participants will be used to track teacher participation (and turnover), students' classroom placements, and data collection completion. Student rosters will be collected prior to the start of each school year. We will track the out-migration and in-migration of each cohort of students based on school attendance data from the time of randomization throughout the implementation years.

Data Analysis Methods. To estimate the impact of *ColSci* on student achievements, we will conduct an intent-to-treat (ITT) analysis using a three-level hierarchical linear model (HLM) with students nested in classrooms and schools (RQ1-3). Schools and their teachers and students will remain in their assigned initial groups. "In-movers," i.e., students who entered the school after random assignment, will be excluded from the analytic sample to avoid the threat of joiner bias (WWC, 2022). Analyses will test the impact of the intervention for 4^{th}-grade students at the end of implementation years one and two, as well as the interventions' differential effects on schools and students with different characteristics, particularly English Learners (ELs). To improve the precision of the impact estimate, each model will adjust for blocking of randomization, baseline measures, and other characteristics at the student, class, and school levels. For the moderator/subgroup analysis, an interaction term of treatment indicator by subgroup will be added to the impact model (RQ4). A two-level regression model will be used to examine the impact of *ColSci* on teacher practice (RQ5), with blocking variables and some teacher-level characteristics as covariates. A two-level structural equation model (SEM) will examine how teacher science practice mediates students' outcomes (RQ6). We will use descriptive and qualitative data analysis to capture how each component is implemented and what factors may hinder or facilitate the implementation (RQ7-8). We will aggregate the data according to the implementation fidelity matrix developed in the first year to determine the level of fidelity and examine whether the effects differ between teachers and students of high vs. low levels of fidelity (RQ9). The analysis for implementation year two will examine the outcomes of students in treatment teachers' second year of teaching with *ColSci* to test the teacher maturation effect (RQ10). Based on school attendance data, we will retroactively identify the students in the study schools at the time of randomization (spring 2026) to determine the sample. We will make school-level inferences if student-level attrition across grade levels is too high. (See the detailed data analysis plan for all RQs in Appendix J11)

D2. Generation of Guidance About Effective Strategies for Replication

The proposed study is designed to generate insightful guidance for the successful replicating and scaling of *ColSci* broadly in four aspects. First, we will strategically sample and recruit schools representing diverse settings to inform generalizability. We will interview district leaders to document the context of implementation and supportive structures and environment, such as district policies, initiatives and priorities, community culture, EL programs, and science and literacy education challenges.

Second, the evaluation will include analysis to assess to what extent the impact of *ColSci* is moderated by the characteristics of students (e.g., ethnicity, socioeconomic status) or schools (e.g., locale, demographic composition) and to identify settings or populations the intervention is particularly effective or less suited (RQ4). *ColSci* was designed to integrate literacy instruction into science teaching, following the NCEE practice guide (Baker et al., 2014), to support ELs in building skills to read, comprehend, and articulate the meaning of increasingly complex informational texts about science. We will investigate the effect of integrated instruction on literacy performance for all students and the subgroup of ELs (RQ2 and RQ4).

Third, the study builds in continuous **monitoring and analysis of the implementation** of each key component in the logic model (RQ7-9), including 1) teachers' participation in PL, 2) consistency in implementing units in *ColSci*, 3) use of embedded assessment with the provided rubrics and motivational prompts, and 4) use of the dashboard to track student progress. WestEd will collect and examine implementation data on all four aspects and provide quarterly feedback to ALI so they can address deficiencies, adapt PL, and adjust teacher support (see Appendix. J14 Formative Evaluation Activities and Timeline**)**. As comparison schools receive delayed treatment in SY 2028-29, we will continue collecting data to examine the replicability (RQ10).

Lastly, WestEd will conduct a **cost analysis** and calculate a **cost-effectiveness** ratio to provide information on the cost of the resources needed to implement *ColSci* with fidelity and sustain *it*. Implementation costs will be identified and gathered systematically in both groups using the "ingredients method" (Levin et al., 2019). It will include all expenditures on various aspects, such as personnel for support, curriculum materials, and PL training. We will use

CostOut - the CBCSE Cost Tool Kit (Hollands et al., 2015-2019) to facilitate the estimation. The analysis will identify costs associated with each key component, distinguish start-up costs from ongoing costs, and calculate the total cost and per-student cost ([redacted] et al., 2024) (see Appendix J13 Cost and Cost Effectiveness Analysis). With the findings, ALI will identify areas for cost-effectiveness improvement for scalability and sustainability.

D3. Components, Mediators, Outcomes, and Measurable Threshold

Student Outcomes. To measure students' achievement in **science** and **literacy**, we will use the state standardized science and English language arts (ELA) assessment, ACAP Summative, at the end of 4th grade. ACAP is a computer-based, criterion-referenced assessment designed to measure student progress on the Alabama Courses of Study Standards. Additionally, WestEd will develop, pilot, and validate a science assessment as a supplemental measure to gauge students' performance on knowledge-in-use assessments. The assessment will be developed in the first year through a principled, clearly defined process grounded in learning and assessment theory ([redacted] et al., 2019). It will include tasks that require students to use their knowledge of scientific ideas and practices to explain phenomena and solve challenging problems to demonstrate their mastery of NGSS performance expectations, which the state science achievement assessment does not capture. These knowledge-in-use assessments will be administered to all students towards the end of 4th grade in each implementation year. We will also analyze EL students' state ACCESS English language proficiency assessment scores (Alabama Comprehensive Assessment Program, 2022**)**. ACCESS is anchored in the World-Class Instructional Design and Assessment (WIDA) English Language Proficiency (ELP) Standards and administered to Alabama students in Grades K-12, identified as ELs. WestEd will secure agreements with participating districts to obtain demographics, enrollment (2nd – 4th grade), and state test scores in 3rd (math and ELA as covariate) and 4th (ACAP science and ELA and ACCESS as outcomes) grades. We will measure students' **motivation and interest** in learning science, including students' perceived competence, intrinsic value, utility value, and attainment value of learning science, using measures based on the Expectancy-Value Theory (Wigfield & Eccles, 2000). All motivational instruments were validated through prior studies and will be administered to both conditions at the beginning and end of each implementation year (See Appendix J12.4 Student Motivation Survey).

Teacher Outcomes. To measure the nature of teacher **science practice** (RQ5) and **fidelity of implementation** of the curriculum, WestEd will administer a weekly log during each *ColSci* unit to capture the instructional activities in the classrooms and enactment of the curriculum ([redacted] et al., 2022, 2023). Logs are a valid and reliable measure of instruction (Ball et al., 1999; Rowan & Correnti, 2009), and WestEd has extensive experience in the successful use of logs to collect data on elementary science instructional practice with high response rates (Iveland et al., 2021). The log will include measures adapted from the instruction practice scales and curriculum coverage scales used by prior studies of reform-oriented instruction (Hamilton et al., 2003; Le et al., 2006). They include reform practices intended to engage students and promote problem-solving skills ($\alpha = 0.83$) and to facilitate critical thinking ($\alpha = 0.77$) and discussion ($\alpha = 0.74$); the teacher reported several lessons on science topics; minutes spent on group work; and hands-on science activities (continuous) (see Appendix J12.1 Science Instructional Log Sample). We will add questions that explicitly ask about teachers' **English language practices,** such as the use of science vocabulary, meaning-making strategies in reading and firsthand investigations, their formative assessment practices, ways to prompt students' motivation, etc. In addition, WestEd will conduct two in-person observations of a random sample of 40% of intervention teachers and 20% of comparison teachers each year to help establish contrast between intervention and comparison classrooms and how observations correspond with teacher self-reports in the logs. The observations will follow the Reformed Teaching Observation Protocol (Piburn et al., 2000; Sawada et al., 2002; reliability R2= 0.95; See Appendix J12.2 Reformed Teaching Observation Protocol Reliability and Subscales Items) to rate teachers' instruction across three domains: Lesson Design and Implementation, Content, and Classroom Culture. We will train observers to use the scales developed by [redacted] et al. (2023) (See Appendix J12.3 Observation Protocol: Multiple Literacies in Project-Based Learning) to capture PBL pedagogies and teachers' efforts to adapt instruction for EL students. An interview after the observations will probe instructional choices and solicit in-depth information about successes and challenges. Teachers will complete a **background** survey adapted from the Teaching and Learning International Survey (Ainley & Carstens, 2018) to gather information on years of teaching experience, professional development, teaching methods, beliefs and attitudes about teaching, and knowledge about NGSS and PBL.

Fidelity of Implementation. WestEd will measure implementation fidelity using the five-step procedure proposed by Nelson et al. (2012). We will develop an implementation fidelity matrix with indicators, metrics, and measurable thresholds for adequate implementation of each key component, including teachers' completion of the PL, the intensity and continuity of the enactment of the *ColSci* units and PBL approach, and the extent to which teachers make use of embedded assessment, motivation prompts, and the dashboard. Teacher logs, observations, and attendance records of PL will serve as data sources. We will interview ALI regional facilitators to reflect on teachers' engagement in PL. A teacher post-survey will gather data on factors that support or hinder their implementation of *ColSci*, their perceptions of the curriculum and classroom resources, changes in their instructional practices, and their satisfaction with PL.

D4. Providing Performance Feedback Towards Achieving Outcomes

During the project's first two years, performance feedback and periodic assessment of progress will be addressed through formative evaluation that aims to augment MSU's development and continuous improvement of *ColSci*. It is guided by research questions about the process and progress of the program's development. Process-related questions include: *Are the newly developed embedded assessment rubrics and motivational prompts, and PL materials appropriately vetted and achieving their objectives?* Progress-related questions include: *Are project milestones being met as intended? Are the new features usable by the targeted users and feasible to be implemented in authentic classroom settings?* To address the questions, the team will interview collaborators, teacher participants, and co-designers, conduct a comprehensive review of project materials and iterative development plans, and offer periodic feedback during the development phase (see detailed timeline of formative evaluation activities in Appendix J14). During the impact study, WestEd will analyze teacher log data and calculate the quarterly fidelity metrics at the teacher level. We will share the results with MSU and the ALI team to inform customized individual support and continuous improvement.

As the independent evaluator throughout the project, WestEd will monitor and track all activities to ensure all goals stay within their scope and timeline. All parties will participate in virtual, monthly calls to discuss upcoming milestones and review objectives and progress to date on development, implementation, data collection, and other project components and steps, as detailed in the five-year project timeline of objectives, milestones, and responsibilities designated for each year of the project (see Appendix J4).

References

Ainley, J. and R. Carstens (2018), Teaching and learning international survey (TALIS) 2018 conceptual framework. *OECD Education Working Papers*, No. 187, OECD Publishing, Paris.

Alabama Comprehensive Assessment Program (2022). Student Assessment Resource Guide for English Learners. Alabama State Department of Education.

Alabama State Department of Education (2024). 2022-2023 Student Assessment State Data Overview. Retrieved from: https://www.alabamaachieves.org/wp-content/uploads/2023/07/RD_2023710_2022-2023-Student-Assessment-State-Data-Overview_V1.0.pdf

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., Gersten, R., Haymond, K., Kieffer, M. J., Linan-Thompson, S., & Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school (NCEE 2014-4012).* Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website: https://ies.ed.gov/ncee/wwc/PracticeGuide/19

Ball, D. L., Camburn, E., Correnti, R., Phelps, G., & Wallace, R. (1999). *New tools for research on instruction and instructional policy: A web-based teacher log*. Seattle, WA: Center for the Study of Teaching and Policy.

Bathgate, M., & Schunn, C. D. (2017). Factors that deepen or attenuate decline of science utility value during the middle school years. Contemporary Educational Psychology, 49, 215-225.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29* (1), 30–59.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8* (2), 225–246.

Davenport, J.L., Kao, Y. S., Matlen, B. J., & Schneider, S. A. (2019). Cognition research in practice: engineering and evaluating a middle school math curriculum. *The Journal of Experimental Education*, 88 (4), 1–20.

[redacted]

Dong, N., & Maynard, R. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, *6*, 24–67. https://doi.org/10.1080/19345747.2012.673143

Dorph, R., Shields, P., Tiffany-Morales, J., Hartry, A., McCaffrey, T. (2011). *High hopes–Few opportunities: The status of elementary science education in California.* Sacramento, CA: The Center for the Future of Teaching and Learning at WestEd.

[redacted]

[redacted]

Hamilton, L, McCaffrey, D., Stecher, B., Klein, S., Robyn, A., and Bugliari D. (2003). Studying Large-Scale reforms of instructional practice: An example from mathematics and science, *Educational Evaluation and Policy Analysis*, Vol. 25, No. 1, 2003, pp. 1–29.

[redacted]

[redacted]

Hollands, F. M., Hanisch-Cerda, B., Levin, H. M., Belfield, C.R., Menon, A., Shand, R., Pan, Y., Bakir, I., & Cheng, H. (2015-19). *CostOut - the CBCSE cost tool kit.* New York: Teachers College, Columbia University, Center for Benefit-Cost Studies of Education. www.cbcsecosttoolkit.org

[redacted]

Jagers, R. J., Rivas-Drake, D., & Williams, B. (2019). Transformative social and emotional learning (SEL): Toward SEL in service of educational equity and excellence. *Educational Psychologist*, 54(3), 162-184.

Kraft, M. (2020). Interpreting effect sizes of education interventions. Educational Researcher, 49(4), 241–253. https://doi.org/10.3102/0013189X20912798

[redacted]

[redacted]

Ladd, G. W., & Dinella, L. M. (2009). Continuity and change in early school engagement: Predictive of children's achievement trajectories from first to eighth grade? *Journal of Educational Psychology*, 101, 190-206.

Le, V. N., Stecher, B. M., Lockwood, J. R.,Hamilton, L. S., & Robyn, A. (2006). *Improving mathematics and science education: A longitudinal investigation of the relationship between reform-oriented instruction and student achievement.* Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/monographs/MG480.html

Lee, O., Deaktor, R., Enders, C., & Lambert, J. (2008). Impact of a multiyear professional development intervention on science achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching, 45*(6), 726–747.

Lee, O., Grapin, S., & Haas, A. (2018). How the NGSS science instructional shifts and language instructional shifts support each other for English learners: Talk in the science classroom. In *Language, Literacy, and Learning in the STEM Disciplines* (pp. 35–52). Routledge.

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2019). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis.* SAGE publications.

McClure, E. R., Guernsey, L., Clements, D. H., Bales, S. N., Nichols, J., Kendall-Taylor, N., & Levine, M. H. (2017). *STEM starts early: Grounding science, technology, engineering, and math education in early childhood.* New York: The Joan Ganz Cooney Center at Sesame Workshop.

[redacted]

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*(1), 18–35.

National Academies of Sciences, Engineering, & Medicine. (2022). Science and engineering in preschool through elementary grades: The brilliance of children and the strengths of educators. The National Academies Press. https://doi.org/doi:10.17226/26215

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.

National Research Council. (2014). *Literacy for science: Exploring the intersection of the Next Generation Science Standards and Common Core for ELA standards: A workshop summary.* National Academies Press.

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., &Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research, 39(*4), 374-396.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.

Palincsar, A. S., & Magnusson, S. J. (2001). The interplay of firsthand and second-hand investigations to model and support the development of scientific knowledge and reasoning. In S. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 151–191). Mahwah, NJ: Erlbaum.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science*, *328*(5977), 459–463.

Piburn, M., Sawada, D., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000) *Reformed teaching observation protocol (RTOP)* (ACEPT Report no. IN-003). Retrieved from https://mathed.asu.edu/instruments/rtop/RTOP_Reference_Manual.pdf

Romance, N. R., & Vitale, M. R. (2001). Implementing an in-depth expanded science model in elementary schools: Multi-year findings, research issues, and policy implications. *International Journal of Science Education, 23*(4), 373–404.

Rowan, B., & Correnti, R. (2009). Studying reading instruction with teacher logs: Lessons from the study of instructional improvement. *Educational Researcher* , 38 (2), 120-131.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., . . . Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics*, *102*(6), 245–253.

[redacted]

Strobel, J., & Van Barneveld, A. (2009). When is PBL more effective? A meta-synthesis of meta-analyses comparing PBL to conventional classrooms. *Interdisciplinary journal of problem-based learning*, 3(1), 44-58.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models.

Trygstad, P.J., Malzahn, K.A., Banilower, E.R., Plumley, C.L., and Bruce, A.D. (2020). *Are all students getting equal access to high-quality science education?* Data from the 2108 NSSME. Horizon Research, Inc.

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.

What Works Clearinghouse (WWC). (2022). *What Works Clearinghouse: Procedures and Standards Handbook, Version 5.0.* Washington, DC: Author. Retrieved from: https://ies.ed.gov/ncee/WWC/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf

White, I. R., & Thompson, S. G. (2005). Adjusting for partially missing baseline measurements in randomized trials. *Statistics in Medicine*, *24*(7), 993–1007.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, *25*(1), 68-81.