# SCALING AND SUSTAINING THE SCIENCE TEACHERS LEARNING FROM LESSON ANALYSIS (STeLLA) PROFESSIONAL DEVELOPMENT PROGRAM

## CONTENTS

## SCALING AND SUSTAINING THE SCIENCE TEACHERS LEARNING FROM LESSON ANALYSIS (STeLLA) PROFESSIONAL DEVELOPMENT PROGRAM

## INTRODUCTION

The growing consensus about the importance of STEM education has put a spotlight on the teaching of science in the elementary grades. Many believe success in elementary science provides a launchpad for future engagement and achievement across STEM areas. But elementary educators lack the supports needed to help their student attain modern standards for science learning. To address this need, BSCS Science Learning and its partners propose a mid-phase grant to scale, refine, test, and sustain a teacher professional development (PD) program proven to improve student achievement in science in elementary grades.

The Science Teachers Learning from Lesson Analysis (STeLLA) program builds on 15 years of research and development and uses a one-year cycle of activities that align to the vision for teaching and learning embodied in the Next Generation Science Standards (NGSS). STeLLA recently demonstrated impacts on teacher outcomes and student science achievement, above and beyond any impacts from a traditional PD program in science (Taylor, Roth, Wilson, Stuhlsatz, and Tipton, 2017; Roth, Wilson, Taylor, Stuhlsatz, and Hvidsten, 2018). The proposal thus addresses *AP1: Moderate Evidence* and *AP3: Promoting STEM Education.*

The project will implement STELLA using a new, scalable delivery model in diverse contexts. The implementation partners include 19 Local Education Agencies (LEAs) in Kentucky and Tennessee, where a significant number of students are high-needs: more than half of potential schools in the participating districts have a rural designation; some are urban; and more than 90% are Title 1 schools. In addition, the partners include regional leaders at PIMSER (the Eastern Kentucky University Partnership Institute for Math and Science Education Reform), Tennessee Aquarium, and Instruction Partners (a Tennessee-based non-profit supporting schools and districts in reform efforts) who will engage state and local leaders, build local capacity, and support dissemination and long-term sustainability.

To facilitate iterative testing and refinement of the scalable delivery model, BSCS and local partners will implement the STeLLA program with fourth and fifth grade teachers across three successive cohorts: Cohort 1 in 2020–21, Cohort 2 in 2021–22, and Cohort 3 in 2022–23, as shown in Exhibit 1. Across these cohorts, BSCS will use feedback routines, internal data, and evaluation data from the project's evaluator—American Institutes for Research (AIR)—to continuously improve its materials and processes for future implementations.

AIR will also conduct an impact evaluation, focusing on Cohorts 2 and 3. AIR will randomly assign schools who want to participate in the program—half to the STeLLA treatment condition and half to receive the opportunity to participate in STeLLA in Year 5. Cohorts 2 and 3 combined will be large enough to provide a sufficiently-powered test (50 schools) of impacts on student achievement.

**Exhibit 1. Number of Teachers in Each Cohort by Year**

| | **Year 1: 2019-2020** | **Year 2: 2020-2021** | **Year 3: 2021-2022** | **Year 4: 2022-2023** | **Year 5: 2023-2024** |
|---|---|---|---|---|---|
| Cohort 1: Pilot | | 24 Teachers | | | |
| Cohort 2 | | | 20 Schools, 60 Teachers | | |
| Cohort 3 | | | | 30 Schools, 90 Teachers | |

## A. SIGNIFICANCE

### A.1. Educational Problem and Effective Strategy

This project addresses science proficiency—a problem for the nation overall and especially for high-need students—and focuses on early grades where challenges first emerge. Policy makers and educators assert that science knowledge is critical to economic productivity and well-being in the U.S., but the National Assessment of Educational Progress (NAEP) shows that 24 percent of fourth grade students fail to reach the “Basic” science achievement level (NAEP, 2015). The problem is most pronounced for black students and Hispanic students (46% and 38%, respectively, compared to 12% of white students) and for students from low-income households (37% compared to 10% from high income households, as measured by eligibility for the National School Lunch Program (NSLP)). These achievement gaps persist and expand as students move into eighth grade, where, for example, 59% of black and 48% Hispanic students fail to perform at the *Basic* achievement level, compared to 18% of white students, pointing to the critical role of an effective science education in early grades.

Frustration with these problems led educators and policy makers to forge a new vision for science teaching and learning—one that addresses past criticism that science instruction in the U.S. lacks focus, connection to student experience, and an authentic view of the scientific enterprise (e.g., NRC, 2007). The development and adoption of research-based teaching practices, represented by the K-12 Framework for Science Education and the Next Generation Science Standards (NGSS Lead States, 2013; NRC, 2012) represents an attempt to transform science instruction and student engagement in science learning. **But most teachers are unprepared to make the dramatic departure from current practices required by contemporary visions of science teaching and learning** (Osborne, 2014; Wilson, 2013; Darling-Hammond, Hyler, Gardner, 2017).

This problem demands effective professional development, but most PD efforts seldom provide teachers with the science content knowledge and pedagogical skills necessary to help them teach in ways called for in current reforms (Reiser, 2013; Wilson, 2013; Darling-

Hammond, Hyler, Gardner, 2017). **The challenges are especially prevalent for elementary teachers** who have little training in science-specific pedagogy or in the science disciplines they are expected to teach (Dorph et al., 2007, 2011; Fulp, 2002; Darling-Hammond, Hyler, Gardner, 2017). In addition, elementary teachers are responsible for teaching multiple subjects and for attending to their professional growth in each of these curricular areas, leaving less time for science-specific PD. **Rural areas face further challenges providing effective teacher learning experiences**, where schools often have small numbers of teachers who can share expertise, and where travel costs to PD sessions can be prohibitive. Rural districts therefore are particularly interested in professional development programs that can be delivered online.

The STeLLA program is an effective strategy that addresses this critical national problem. STeLLA engages teachers in sustained (year-long) analysis of science teaching and learning in professional learning communities. The work of the communities is grounded in a conceptual framework focused on *student thinking* and *instructional coherence* that stimulates professional learning and supports teachers in designing and delivering more effective science instruction, leading to improved classroom practice and student achievement.

### A.2. An Unmet Demand: Scalable Science PD with Demonstrated Effectiveness

The demand for effective programs like STeLLA is strong. Through programs such as ESEA Title II, school districts have funds available to increase student achievement by recruiting, training, and retaining high quality teachers, particularly in high-needs local educational agencies. Further, the rapid adoption of the Next Generation Science Standards (NGSS) means districts and states are looking for proven approaches to support teacher implementation of this new vision of science teaching and learning. So far, 19 states and the District of Columbia (representing over 36% of U.S. students) have adopted the Next Generation Science Standards

(NGSS), while 21 states (representing 33% of U.S. students) have developed their own standards based on recommendations in the NRC Framework for the NGSS. Each of these states is in the process of developing and piloting assessments aligned with the NGSS, resulting in strong demand for effective science teacher professional development.

The adoption of the Every Student Succeeds Act (ESSA) further expands demand for programs like STeLLA by requiring districts to use funds on evidence-based practices. In science education very few studies meet the standards of research that enable us to make causal claims about the impact of PD on student learning (Hill, Beisiegel, & Jacob, 2013; Little, 2011; Wilson, 2011, 2013; Yoon et al., 2007; Darling-Hammond, Hyler, Gardner, 2017). This lack of evidence on effective models of PD limits our ability to support teachers in aligning their teaching practice with reforms such as the NGSS (NGSS, 2013; Reiser, 2013). Evidence on how to scale up effective PD programs is also scarce, with studies noting multiple challenges in taking programs to scale while maintaining effectiveness (Kraft, Blazar, & Hogan, 2018).

**STeLLA is one of a handful of science professional development programs with demonstrated evidence of effectiveness on teacher and student outcomes** (Roth et al., 2011; Taylor, Roth, Wilson, Stuhlsatz, and Tipton, 2017; Roth, Wilson, Taylor, Stuhlsatz, and Hvidsten, 2018; Kennedy 2016; Darling-Hammond, Hyler, Gardner, 2018). In a cluster randomized trial (CRT) with 140 teacher participants in Colorado, we estimated the effects of the STeLLA PD program relative to the same amount of PD focused solely on deepening teachers' science content knowledge, exploring the pathways of influence shown in Exhibit 2.

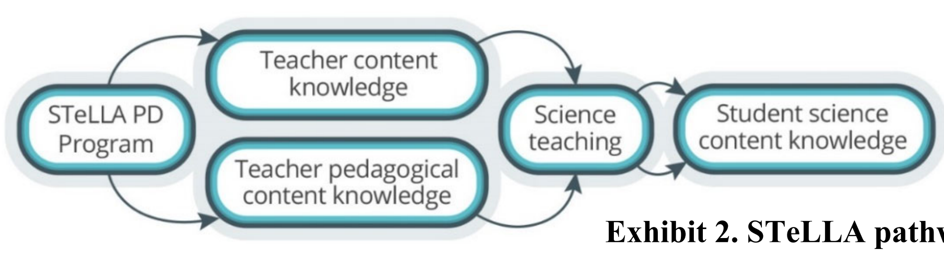



**Exhibit 2. STeLLA pathway of influence**

Teachers were randomly assigned by school to either the STeLLA program or the comparison Content Deepening program. Both groups received the same number of hours of PD. The effect of STeLLA was noteworthy with a corresponding effect size of 0.68 ($p<.001$) on a measure of students' science achievement (Taylor, Roth, Wilson, Stuhlsatz, and Tipton, 2017). At the teacher level, teacher's science content knowledge (effect size = 0.66), pedagogical content knowledge (effect size = 1.17), and classroom practice (effect size = 2.05) were all significantly higher ($p<.05$) in the STeLLA program than in the Content Deepening program (Roth, Wilson, Taylor, Stuhlsatz, & Hvidsten, 2018). This study provides strong evidence of effectiveness by meeting What Works Clearinghouse (WWC) Evidence Standards without reservations.

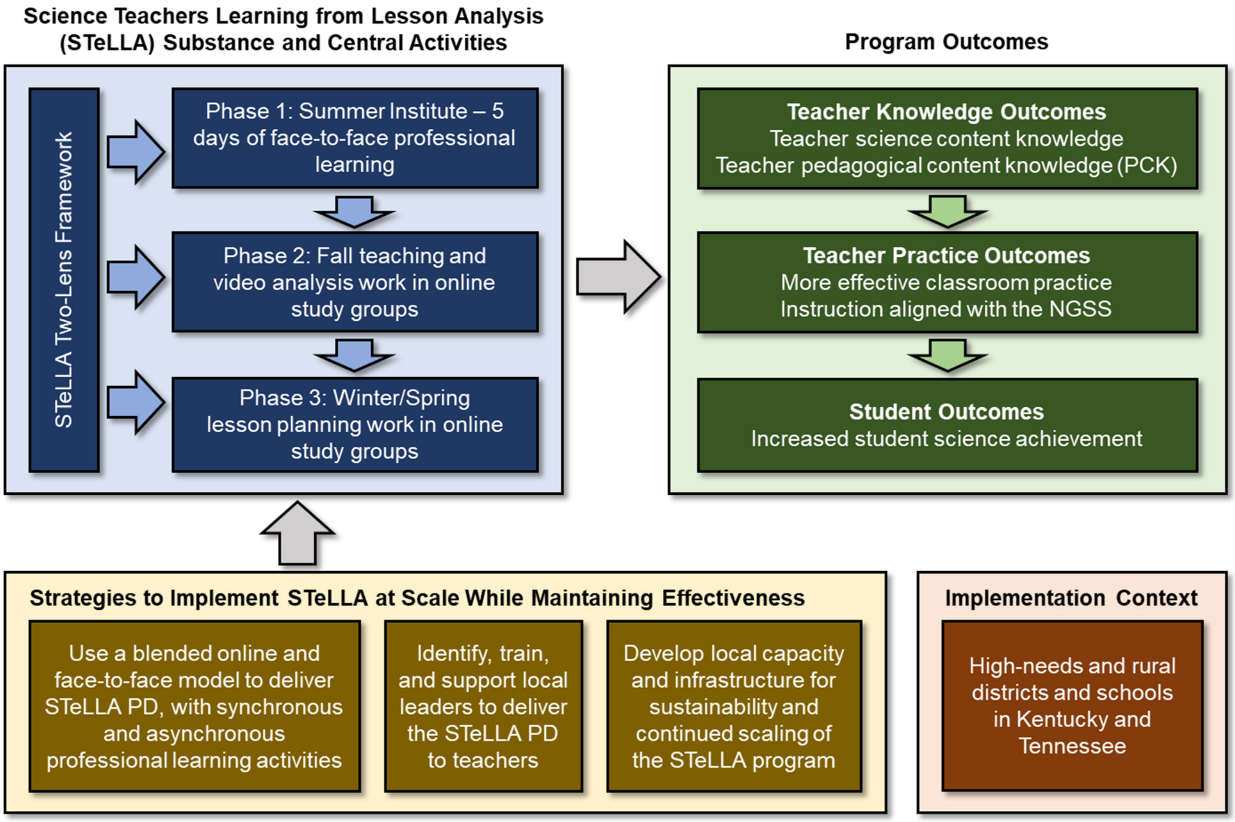
As such, **STeLLA is proven to have impacts, even above what one can expect from a similarly-intensive but traditional science PD program.** The demonstrated effectiveness of the STeLLA model and additional factors discussed above led 19 districts to submit letters of interest. Given the strong statistical findings and positive participant experiences in previous STeLLA studies, along with the continued spread and adoption of the NGSS, we expect demand for the STeLLA program to continue to rise. Taking the STeLLA program online will further increase demand, as the program becomes more conducive to addressing the constraints of rural districts. **Given the magnitude of the educational problem, the unmet demand for effective and scalable science professional development, and the strong evidence base for the STeLLA program, the challenge now is to scale the program without losing effectiveness.**

## B. QUALITY OF THE PROJECT DESIGN

### B.1. Conceptual Framework

Using the scaling supports described in section C1, professional development leaders in each district will implement the STeLLA program with the participating teachers, leading to impacts on teacher and student outcomes, as depicted in the conceptual framework (Exhibit 3). In this section, we describe the STeLLA program in detail. Measures of implementation fidelity, teacher outcomes, and student outcomes are described in the Project Evaluation in Section E.

**Exhibit 3. STeLLA Conceptual Framework.**




The STeLLA PD program is unique in weaving together a rich, focused, and sequenced set of analysis-of-practice tasks for teachers that are both intellectually challenging and highly practical in supporting teacher change, as described next in the subsections on the program's (a) Substance and (b) Form and Central Activities. These features of the STeLLA PD program are grounded in a growing consensus that professional development should a) engage teachers actively in collaborative analyses of their practice; b) treat content as central and intertwined with pedagogical issues; c) enable teachers to see these issues as embedded in real classroom contexts; d) focus on the content and curriculum teachers are teaching; and e) be guided by an articulated model of teacher learning that specifies what knowledge and skills teachers will gain, what activities will lead to this learning, and how these new skills and knowledge will appear in their teaching practices (Ball & Cohen, 1999; Darling-Hammond & Sykes, 1999; Desimone, 2009; Elmore, 2002; Garet, Porter, Desimone, Birman, & Yoon, 2001; Guskey & Yoon, 2009; Hawley & Valli, 1999).

**STeLLA PD Program: S*ubstance***

The STeLLA program substance focuses on three types of knowledge and abilities that represent the learning goals for teachers.

***1. Teacher Science Content Knowledge.*** Teachers at each grade level are supported in deepening their understandings about a set of core science ideas in two topic areas. Work on developing these science content understandings occurs in the context of analyzing lesson videos and student work across the entire program as well as in content deepening activities led by content experts. Teachers are also supported by a document that describes science content ideas in relevant teaching contexts such as how common student ideas are related to science ideas, how commonly used lesson activities develop or fail to develop core science ideas, or how content representations and analogies have strengths and weaknesses in terms of scientific accuracy and in terms of supporting student learning.

***2. Teacher Science Pedagogical Content Knowledge (PCK).*** Following the work of Shulman (1987), the STeLLA development team drew from a broad literature review to select two types of PCK that were hypothesized to have the most potential to impact teacher and student learning within a one-year time frame: (a) knowledge about students' thinking and difficulties they encounter in learning about specific science content, and (b) knowledge about how to link science ideas to each other and to classroom activities to create a coherent science content storyline (Roth et al., 2017). These aspects of PCK are embodied in the STeLLA Two-Lens framework (Exhibit 4), which provides teachers with eight teaching strategies to make student thinking more visible and nine strategies to support the development of coherent science content storylines.

***3. Ability to Use Content and Pedagogical Content Knowledge in Planning and Teaching Science.*** We assume that content knowledge and pedagogical content knowledge about science content storylines and student thinking will only impact student learning if teachers apply that knowledge in their teaching. Teachers are therefore supported in translating that knowledge into effective teaching through model lesson plans that highlight the STeLLA lenses and strategies, videos from external teachers' classrooms, lesson planning tools, and collaborative structures.

**Exhibit 4. The STeLLA Two-Lens Framework**

| Student Thinking | Science Content Storyline |
|---|---|
| **Student Thinking Lens**: Strategies to Reveal, Support, and Challenge Student Thinking | **Science Content Storyline Lens**: Strategies to Create Coherent Science Content Storylines |
| 1. Ask questions to elicit student ideas and predictions<br>2. Ask questions to probe student ideas and predictions<br>3. Ask questions to challenge student thinking<br>4. Engage students in analyzing and interpreting data and observations<br>5. Engage students in constructing explanations and arguments<br>6. Engage students in using and applying new science ideas in a variety of ways and contexts<br>7. Engage students in making connections by synthesizing and summarizing key science ideas<br>8. Engage students in communicating in scientific ways | A. Identify one main learning goal<br>B. Set the purpose with a focus question or goal statement<br>C. Select activities that are matched to the learning goal<br>D. Select content representations and models matched to the learning goal and engage students in their use<br>E. Sequence key science ideas and activities appropriately<br>F. Make explicit links between science ideas and activities<br>G. Link science ideas to other science ideas<br>H. Highlight key science ideas and focus question throughout<br>I. Summarize key science ideas |

**STeLLA PD Program: *Form and Central Activities***

Central to the form of the STeLLA program is teachers' interactions with each other and with videocases as they analyze science content, science teaching, and science learning in small (5-8 teachers) grade-level study groups, each led by a PD leader. These interactions begin during a five-day summer institute (40 hours) and continue in monthly 3 to 5-hour meetings across the school year (30 hours) and asynchronous work (20 hours). In total, there are 90 hours of PD in the one-year program. There are three basic phases of the PD work:

***Phase 1: Summer Institute.*** The program begins with 5 days of face-to-face PD to build trust and norms of collaboration within the learning community, and to investigate the teacher science content learning goals of the program. Teachers are introduced to the STeLLA lenses and strategies, and the lesson analysis process. They analyze video clips of experienced teachers and practice identifying the STeLLA strategies in use in these classrooms.

***Phase 2: Fall Teaching and Video Analysis Work in Online Study Groups, Content Area 1.*** During the fall, teachers teach the lesson provided by the program that are used as model lessons for the teachers. In monthly 3 to 5-hour online study group meetings they learn from analysis of videos from each other's teaching of these lessons and from analysis of student work. Outside of the synchronous study group sessions, teachers participate in asynchronous program experiences, analyzing classroom videos to enhance understanding of science content, student ideas about the content, and strategies to respond to student ideas and move student thinking forward.

***Phase 3: Winter/Spring Lesson Planning Work in Study Groups, Content Area 2.*** The third phase of the STeLLA program begins in December when each study group switches to a second topic area. In addition to switching content areas, the focus of the monthly study group sessions shifts from implementing provided model lesson plans to developing their own lesson plans, using the STeLLA lenses and strategies as guides. The program ends with the teachers teaching the lessons they have developed and meeting for a final time to share analyses of this teaching experience.

These three phases are designed to build on each other. Consistent with the situated cognition theory of teacher learning (Brown, Collins, & Duguid, 1989) and the cognitive apprenticeship instructional model (Lave, 1988). What varies over time are the amount and type of scaffolding by the PD leader, the nature of the teachers' activities, and the nature and role of the videocases. The videocases, introduced in Phase 2, are perhaps the most critical ingredient. **Each videocase is a compilation of video clips, lesson plans, and student work generated in classrooms outside the treatment group.** The videocases highlight science content, science teaching, and science learning, and provide the context for teachers to practice paying attention to the science ideas and misconceptions students reveal during instruction.

**Implementation Context.** To examine the scaling of the STeLLA program with high needs students, we situate this study in Kentucky and Tennessee – a region with a significant number of students at risk for educational failure. Rural districts in the region face unique challenges that hinder improvement efforts, including high poverty and dropout rates, difficulties finding highly qualified teachers, and a lack of available resources. While some non-rural districts have expressed interest, more than half of the teachers will be from rural schools. The Technical Appendix shows district data from which we have obtained **letters of commitment**.

## B.2. Clearly Specified and Measurable Goals, Objectives, and Outcomes

The project's goal is to test and refine a strategy for scaling STeLLA in diverse settings that serve high-need students, and to build a network to support continued scaling. Exhibit 5 specifies the objectives, strategies, and outcomes to be achieved and how outcomes will be measured.

**Exhibit 5. Objectives, Strategies, Outcomes, and Measures.**

| Strategies | Outcomes | Measures |
|---|---|---|
| **Objective 1: Adapt the STeLLA program to a blended online and face-to-face format, using continuous feedback and implementation data.** | | |
| Strategy 1.1. Transfer current STeLLA PD materials to the Canvas online platform. | Online portal for STeLLA materials, including strategy guides, lesson plans, assessments, student work, and classroom video. | Measure 1.1. STeLLA website. Memos documenting online materials development. 100% complete by end of Year 1. |
| Strategy 1.2. Pilot blended PD model with 24 teachers in Year 2 | Participant feedback and implementation data on STeLLA materials | Measure 1.2. PD observations; At least 90% attendance at PD sessions; classroom videos; teacher surveys. |
| Strategy 1.3. Refine materials and procedures for each strategy under Objective 1. | Improved STeLLA materials, PD leader guides, STeLLA strategy guide | Measure 1.3. Bi-annual memo documenting revisions made to the STeLLA materials and resources. Objective 1 contains all required elements. |
| **Objective 2: Develop local leaders to deliver the STeLLA program and revise leadership program using continuous feedback and implementation data.** | | |
| Strategy 2.1. Work with local partners to identify potential PD leaders. | Pool of potential PD leaders and PD leader selection criteria. | Measure 2.1. Selection criteria; spreadsheet documenting qualifications of potential leaders. |
| Strategy 2.2. Select and recruit local staff to serve as PD leaders. | Well-qualified leaders committed to learning and implementing STeLLA PD with fidelity. | Measure 2.2. Leader applications; applicant screening and interview records; 100% of new leaders hold minimum qualifications or better. |
| Strategy 2.3. Prepare local staff to serve as STeLLA PD leaders. | Fully trained leaders for each study group. | Measure 2.3. New leaders attend >90% of training; training session observations; leader interviews |
| Strategy 2.4. Support local leaders in implementing STeLLA PD. | Leaders receive individualized feedback and support from BSCS expert STeLLA leaders. | Measure 2.4. Fidelity of implementation observations of >50% of PD sessions; teacher participant surveys; PDL interviews; attendance at leader mentoring sessions. |
| Strategy 2.5. Refine materials and procedures for each strategy under Objective 2. | Improved leadership development program and resources. | Measure 2.5. Biannual memo documenting revisions made to the leadership program. Objective 2 contains all required elements. |
| **Objective 3: Implement the blended STeLLA program while continuously using feedback and fidelity data for program improvement.** | | |

| | | |
|---|---|---|
| Strategy 3.1. Local leaders implement STeLLA orientation in each district. | Treatment teachers, PD leaders, principals and district points of contact understand STeLLA and commit to participation. | Measure 3.1. 100% attendance at orientation sessions; teacher orientation form. |
| Strategy 3.2. Local leaders implement face-to-face Summer Institutes at regional hub locations. | Teachers learn STeLLA Student Thinking lens and Content Storyline lens strategies. Collective analysis of classroom video. | Measure 3.2. 90% of teachers have one or fewer absences at PD sessions. Teacher content knowledge and PCK pretests; PD session observations. |
| Strategy 3.3. Local leaders implement monthly study group meetings via online Zoom meetings. | Collective analysis of classroom video, teachers implement STeLLA strategies and STeLLA lessons, leading to improved classroom practice and student achievement. | Measure 3.3. 90% of teachers have one or fewer absences at PD sessions; recordings of 100% of online PD sessions; teacher feedback surveys; classroom video. |
| Strategy 3.4. Refine materials and procedures for each strategy under Objective 3. | Improved STeLLA PD program and associated resources. | Measure 3.4. Biannual memo documenting revisions made to the STeLLA materials and resources. Objective 3 contains all required elements. |
| **Objective 4: Conduct a CRT to test the impact of the STeLLA program on teacher and student outcomes.** | | |
| Strategy 4.1. Identify and recruit eligible teachers in participating schools. | Eligible 4th and 5th grade teachers interested in STeLLA and who consent to random assignment at the school level. | Measure 4.1. Recruit 20 schools for cohort 2 and 30 for cohort 3. Consent forms from 100% of teachers and principals. |
| Strategy 4.2. Randomly assign schools to treatment and control conditions. | Samples of treatment and control schools with likely baseline equivalence. | Measure 4.2. Random assignment memo; baseline equivalence memo. |
| Strategy 4.3. Measure and analyze STeLLA PD implementation fidelity. | Fidelity of implementation data, collected and analyzed. | Measure 4.3. Classroom observations; teacher implementation surveys. Response rate of 90% or more. |
| Strategy 4.4. Measure and analyze treatment-control contrast in teachers' PD experiences. | Data on participation in PD in treatment and control groups, collected and analyzed. | Measure 4.4. Data collection updates, report on treatment-control contrast. |
| Strategy 4.5. Determine the impact of STeLLA on teacher knowledge, | Data on outcome measures, collected and analyzed. | Measure 4.5. Teacher content knowledge, PCK and practice data. Student achievement data from |

| | | |
|---|---|---|
| classroom practice, and student achievement. | | >80% of students. Report that meets WWC standards without reservation. |
| Strategy 4.6. Refine materials and procedures for each strategy under Objective 4. | Improved procedures for teacher recruitment, random assignment, data collection, and analysis. | Measure 4.6. Biannual memo documenting revisions made to materials and procedures for Objective 4. Objective 4 contains all required elements. |
| **Objective 5: Develop capacity and infrastructure for sustainability and continued scaling of the STeLLA program.** | | |
| Strategy 5.1. Establish and operate network to support and sustain work of participating districts. | Regional network of district leaders, PD leaders and teachers committed to supporting and sustaining the STeLLA program. | Measure 5.1. Roster of network members; network website; network meeting materials. |
| Strategy 5.2. Support local leaders to identify new locations and teachers who would benefit from STeLLA. | Expanding pool of STeLLA teachers and STeLLA teacher leaders | Measure 5.2. Signed MOUs from additional interested districts. |
| Strategy 5.3. Develop and refine approach to help districts extend STeLLA into existing PD systems. | Resources and guides to support the integration of STeLLA PD into district teacher improvement systems. | Measure 5.3. Committee agenda; participation from 100% of partners, summary of committee input. Objective 5 contains all required elements. |

## C. STRATEGY TO SCALE

### C.1. Specific Strategies to Implement STeLLA at Scale with High Quality

**Strategy 1. Move from face-to-face to online delivery, with synchronous and asynchronous professional learning experiences.**

The project will implement several specific strategies to scale that are intended to make the program more flexible for participants, facilitate high-quality implementation at scale, and build infrastructure to sustain the program beyond the life of the grant.

Professional development that relies on face-to-face delivery presents a significant barrier to scaling, as it requires considerable travel from both teachers and PD leaders, scheduling at times when all participants are available, and limits the number of teachers a PD leader can support. Our strategy to addressing this barrier is the implement a **blended face-to-face and**

**online version of the STeLLA program**. This blended approach replicates the STeLLA face-to-face program with the same learning goals, key program features, and total program hours, and takes advantage of online delivery to allow for flexibility of location (especially for rural teachers), variability in the timing of individual participation, and lower costs of delivery and participation. The online delivery will draw on emerging design principles for online and blended PD in STEM education, including motivating and sustaining engagement that builds knowledge and advances professional goals; creating opportunities for teachers to collaborate as learners; and supporting reflection on content and practice (CADRE, 2017). In prior work, BSCS has developed and studied an online version of the STeLLA program for use in pre-service elementary science courses which showed positive impacts on both early career teachers and their students (Wilson, Stuhlsatz, Hvidsten, & Stennett, 2017). We are also working with funding from the Minnesota Department of Education to introduce a portion of the STeLLA program online to teachers across that state (MSP #2016-00170).

**Strategy 2. Develop local leaders to deliver STeLLA PD**

A significant barrier to scaling STeLLA in the past is the program's reliance on BSCS staff to deliver the program. This approach is difficult to scale due to the limited number of BSCS staff available, geographic constraints, and the desire from many districts and teachers to receive PD from trusted local sources who understand the local context. To address this barrier, we will implement a local STeLLA PD leader development program where BSCS leadership development experts prepare and support local PD leaders, who then lead STeLLA PD for teachers. During the first year of the leadership program, local leaders will complete a summer Leadership Institute and Learning Team meetings during the academic year. The first year of the leadership program has a dual focus. Participants first experience the STeLLA PD program as learners with a focus on the program experiences, and then shift focus to the knowledge and skills needed to lead the program. The new leaders are then mentored by experienced BSCS staff and implement the STELLA program with their own groups of teachers (Cohort 1, Pilot). As shown in Exhibit 5 (Section B.2), the leadership development program involves three key stages – identification and recruitment of local leaders, leadership development, and leader monitoring and support. Monitoring and support will be provided by BSCS staff, informed by observations and data from the online PD delivery system, which will help BSCS continuously improve the leadership development program. BSCS has extensive leadership development experience. Since 2013, PD leaders throughout the country have participated in BSCS Professional Development Providers Institutes to learn the knowledge and skills necessary for NGSS-aligned PD design and facilitation. In developing the STeLLA Leadership program, BSCS will refine existing leadership development models that were used in Minneapolis Public Schools, St. Paul Public Schools, Jefferson County Schools in Louisville, KY, and the Pomona Unified School District.

**Strategy 3. Develop capacity for sustainability and continued scaling of the program.** With this third strategy, we recognize that many professional development programs supported with federal funds end with the completion of the award. To address this barrier, we will:

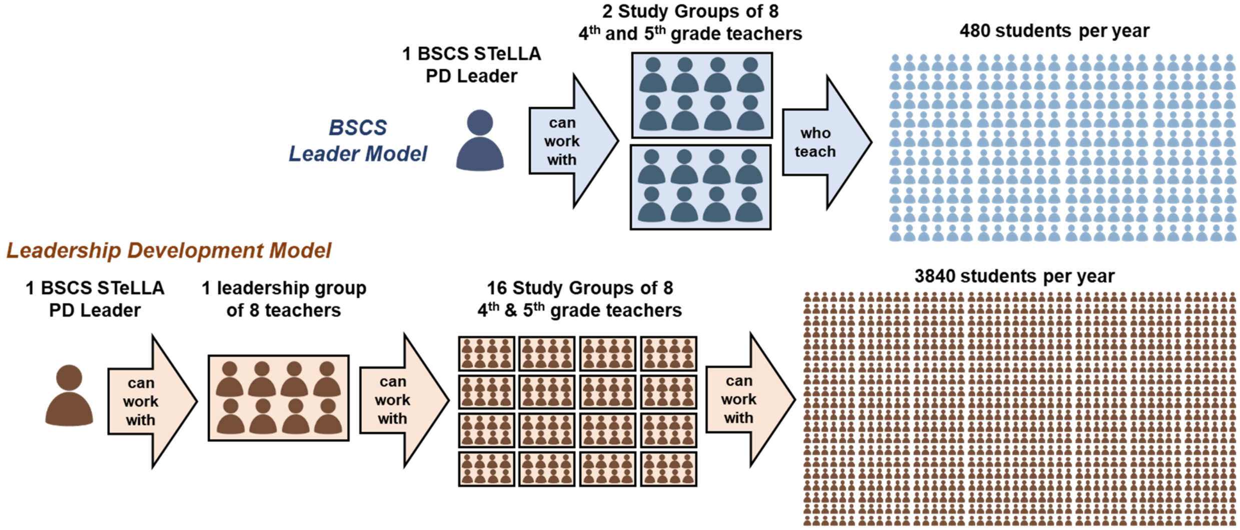
a) Work with local organizations and PD leaders to provide STeLLA professional learning experiences to teachers in treatment schools who were not able to participate in the study.
b) Work with local organizations and PD leaders to provide STeLLA professional learning experiences in year 5 to those teachers who were assigned to the control group.
c) Provide ongoing free access to all materials associated with the project across the region, including lesson plans, PD leader guides, classroom video, and video analysis tools.
d) Convene bi-annual meetings across all partners to develop long term sustainability plans.

### C.2. Increasing Efficiency

Scaling the program demands that it become more efficient and cost-effective, while maintaining its transformative impacts on teaching and learning. To this end, each of our scaling strategies are designed to increase the efficiency of the program. In Strategy 1, delivering the STeLLA program across regions and to teachers in rural and isolated areas costs less by reducing the need for travel, meeting facilities, and support materials. We project a 59% reduction in travel costs and one-third less facilitation time compared to face-to-face delivery by program developers.

The leadership development approach in Strategy 2 provides a significant increase in efficiency over previous models where BSCS staff led all PD sessions. Educating local leaders and school-district personnel to lead PD is a cost-effective and sustainable approach for delivery (Borko, Jacobs, Koellner & Swackhamer, 2015) and is a one-time cost. Once local leaders have the requisite skills and abilities to effectively implement the program, they can do so with different groups of teachers over an extended period. Exhibit 6 shows the efficiency obtained by moving to the leadership development model employed in this study.

**Exhibit 6. Efficiency obtained by moving to a local leadership development model.**




Finally, we will make the program more efficient through continuous improvement and refinement across cohorts, based on implementation and cost data from the evaluation.

## D. ADEQUACY OF RESOURCES AND QUALITY OF THE MANAGEMENT PLAN

### D.1. A Management Plan Defining Responsibilities, Timelines, and Milestones

Exhibit 7 describes the strategies (see Section B.2), the responsible partner, and timing for each milestone in the project's 5-year timeline (see Exhibit 5).

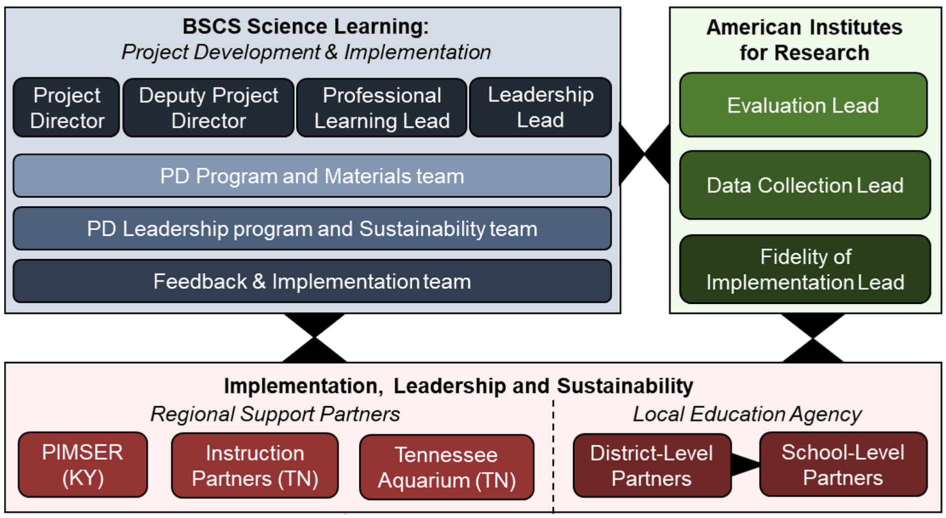
**Exhibit 7. Milestones, group responsible, and timeframe for each strategy.**

| | | Project Year (October 1 to September 30) | | | | |
|---|---|---|---|---|---|---|
| **Milestones** | **Responsible** | **Year 1** | **Year 2** | **Year 3** | **Year 4** | **Year 5** |
| **Objective 1: Revise the STeLLA program to a blended online and face-to-face program, using continuous feedback and implementation data.** | | | | | | |
| Strategy 1.1 | BSCS | √ | | | | |
| Strategy 1.2 | BSCS, PIMSER, IP, TN Aq | | √ | | | |
| Strategy 1.3 | BSCS | √ | √ | √ | √ | √ |
| **Objective 2: Develop local leaders to deliver the STeLLA program and revise leadership program using continuous feedback and implementation data.** | | | | | | |
| Strategy 2.1 | BSCS, PIMSER, IP, TN Aq | √ | | | | |
| Strategy 2.2 | BSCS, PIMSER, IP, TN Aq | √ | √ | | | |
| Strategy 2.3 | BSCS, PIMSER, IP, TN Aq | | √ | √ | √ | |
| Strategy 2.4 | BSCS, PIMSER, IP, TN Aq | | √ | √ | √ | |
| Strategy 2.5 | BSCS | √ | √ | √ | √ | √ |
| **Objective 3: Implement the blended STeLLA program while continuously using feedback and fidelity data for program improvement.** | | | | | | |
| Strategy 3.1 | BSCS, PIMSER, IP, TN Aq | √ | √ | | | |
| Strategy 3.2 | BSCS, PIMSER, IP, TN Aq | | √ | √ | √ | |
| Strategy 3.3 | BSCS, PIMSER, IP, TN Aq | | √ | √ | √ | |
| Strategy 3.4 | BSCS | √ | √ | √ | √ | √ |
| **Objective 4: Conduct a CRT to test the impact of the STeLLA program on teacher and student outcomes.** | | | | | | |
| Strategy 4.1 | BSCS, AIR | √ | √ | | | |
| Strategy 4.2 | AIR | | √ | √ | | |
| Strategy 4.3 | AIR | | √ | √ | √ | |
| Strategy 4.4 | AIR | | √ | √ | √ | |
| Strategy 4.5 | AIR | | √ | √ | √ | |
| Strategy 4.6 | BSCS, AIR | | √ | √ | √ | √ |
| **Objective 5: Develop capacity and infrastructure for sustainability and continued scaling of the STeLLA program.** | | | | | | |
| Strategy 5.1 | BSCS, PIMSER, IP, TN Aq | √ | √ | √ | √ | √ |
| Strategy 5.2 | BSCS, PIMSER, IP, TN Aq | | | √ | √ | √ |
| Strategy 5.3 | BSCS, PIMSER, IP, TN Aq | | | √ | √ | √ |

## D.2. Capacity to Bring the Project to Scale

The management plan determines the reporting relationships for the partner organizations (see Exhibit 8). The plan is more than adequate in part because each partner organization is highly qualified for a clear and specific role involving execution of the strategies (see Section C.1) at each milestone on the project's 5-year timeline (see Exhibit 5).

**Exhibit 8. Project Organization Chart and Reporting Relationships**




BSCS Science Learning is the lead organization for the study responsible to the U.S. Department of Education for grant performance. BSCS Science Learning's primary project roles will be to a) adapt the STeLLA materials to form the blended version, b) recruit and develop local PD leaders, c) support the implementation of the STeLLA PD program by local leaders, d) continuously improve project materials and resources, and e) oversee the agreements with partner organizations (AIR, PIMSER, Instruction Partners, TN Aquarium). BSCS is in its 61st year as a leader in science education, with a commitment to transforming science teaching and learning for all students. BSCS has demonstrated its capacity for developing, implementing, and studying innovative science professional development through IES and NSF awards, with recent publications from those studies in the *Journal of Research on Educational Effectiveness*, the *American Educational Research Journal*, and the *Journal for Research on Science Teaching*.

BSCS staff have the expertise required to develop and implement effective PD programs, as well as the experience managing complex projects required to anticipate and prevent issues before they arise. BSCS has been developing, revising, and researching the STeLLA PD program for 15 years, and will ensure implementation is of high quality and consistent across sites. In addition to developing the line of research that has built the evidence base to take STeLLA to scale, BSCS is uniquely poised to deliver the scaling strategies we propose, as described in C1. ***BSCS Lead Staff:*** Christopher Wilson, Project Director (PD); Molly Stuhlsatz, Deputy Project Director (DPD); Jody Bintz, Leadership and Sustainability Lead; Connie Hvidsten, Professional Learning Lead.

To ensure the independence of the evaluation, **AIR** will conduct all aspects of the evaluation and will have no role in the development or the implementation of the STeLLA intervention except to share implementation analyses as feedback. This structure ensures the independence of key evaluation activities including random assignment, outcome data collection, analysis, and reporting, and is consistent with OII guidance (U.S. Department of Education, Office of Innovation and Improvement, n.d.; Abt Associates, 2015). AIR is uniquely qualified for this role, having successfully led four projects for IES in the last decade focused on teacher PD interventions. These projects involved coordinating across organizations, including an intervention provider and several school districts, recruiting participating schools and teachers, and conducting an independent evaluation. AIR's experience monitoring intervention providers and providing feedback on fidelity for continuous improvement also helps ensure relevant, actionable feedback from the evaluation team, which will draw on instruments and methods that AIR has refined across several studies. ***AIR Lead Staff:*** Joseph Taylor, Evaluation Lead; Seth Brown, Data Collection Lead; Andrew Wayne; Senior Advisor.

**PIMSER**, **Instruction Partners**, and the **Tennessee Aquarium**, will each serve as regional hubs that will a) provide and identify local PD leaders, b) assist in recruiting schools and teachers, c) assist with improvement of the program and materials, and d) support efforts to build regional capacity for sustainability and continued scaling. PIMSER is based at East Kentucky University, and provides support for improvements in STEM education across the state. Instruction Partners in Nashville, TN, partners with districts to support academic improvement, with a focus on rural areas. The Tennessee Aquarium in Chattanooga, TN, provides standards-based educational programs and services across east Tennessee. Multiple **school districts** across KY and TN, particularly in high needs and rural areas, have committed to participating in the study, assisting with teacher recruitment, sharing data on student achievement and demographics, and to contribute resources to support the work.

### D.3. Ongoing Work Beyond the End of the Grant

As a result of this project, our PIMSER, Instruction Partners, and Tennessee Aquarium hubs and school district partners will have a cadre of trained STeLLA leaders, invested school and district leaders, and a cost-effective model for delivering the STeLLA program, which will help ensure the sustainability of STeLLA PD and practices. Local leaders will continue to implement STeLLA PD with new cohorts of teachers and contribute to the expansion of the program in their area. **Moving beyond the regional level, we will offer leadership development program in various locations across the country.** These offerings will attend to STeLLA at grades 4 and 5 as well as at other grade levels.

All PD materials will be made available across the region, including lesson plans, leader guides, classroom video, and the video analysis tools and processes. The online PD environment will continue to be supported and integrated into work with other districts. Finally, BSCS will continue to use all materials in future implementations of the STeLLA program, using data from the AIR evaluation to inform future improvement in outcomes and cost effectiveness. We will also explore the potential for continued scaling beyond the region, particularly via national dissemination through conference presentations and publications.

### D.4. Reasonable Costs / Anticipated Results and Benefits

In the face-to-face STeLLA delivery by the BSCS developers, 12 PD leaders would be needed to prepare the 75 treatment teachers in this study, with a total cost of [redacted] (including organization costs and travel). This results in a cost per STeLLA teacher cost of $[redacted] and a per student cost of $[redacted] assuming 24 students per teacher. **Moving to online delivery of the program with local leaders reduces the total cost to $[redacted] with a per STeLLA teacher cost of $[redacted] and a per student cost of $[redacted]** With the project activities designed to develop capacity for long-term sustainability and expansion of the STeLLA program within the region beyond the term of the project, the cost per student decreases dramatically.

The magnitude of the STeLLA effect is large, and so the anticipated benefits of this investment is significant. In a cluster randomized trial exploring the impact of the STeLLA program, the students of teachers who received STeLLA PD increased the equivalent of 18 percentile points in comparison to students who received conventional PD. This suggests that if the comparison students were at the mean of a normed sample (the 50$^{th}$ percentile) then STeLLA students would be placed at the 68$^{th}$ percentile. **This dramatic increase in science achievement at these early grades will have long lasting impacts on the academic achievement of these students as they move into middle and high school, particularly for high-needs students.** Further, since elementary teachers teach multiple subjects, integrating STeLLA strategies into classroom teaching outside of science expands the impact of this work.

The project is able to offer these benefits at reasonable costs to the government because BSCS has extensive experience developing and managing budgets for large-scale, multi-year projects with multiple partners. Throughout the project we produce monthly reports that track and compare actual project financial performance to budget. We also use flexible staffing models ensure that people with the necessary skills are allocated at the appropriate FTE to appropriate tasks. The budget for this 5-year project has been carefully planned to adequately yet efficiently support all project activities, and has been generated by analyzing actual costs from prior STeLLA projects. The budget includes stipends and travel costs for teacher participants, funds for local leaders and coordinators, as well as costs for dissemination and sustainability activities.

## E. QUALITY OF THE PROJECT EVALUATION

### Overview of the Project Evaluation Design

AIR will lead an independent evaluation of BSCS Science Learning's STeLLA program, as delivered using the project's strategy to scale, including a rigorous impact evaluation, an assessment of implementation fidelity, and an analysis of cost. AIR will conduct a cluster (school) randomized trial (CRT) that will compare outcomes from STeLLA schools to those of Business-as-Usual (BaU) control group schools. The evaluation will include an investigation of how the STeLLA program is implemented at scale, relative to a consistent set of standards for intended STeLLA implementation. AIR also will conduct a cost-effectiveness analysis that will provide critical insights into the investment of resources required for high-fidelity implementation of STeLLA and the association between program cost and program impact on student and teacher outcomes. The evaluation will measure all key program components, mediators, and outcomes described in the program theory of influence (see Exhibit 2).

AIR's evaluation will address the following key research questions (RQs):

**Main impact of STeLLA:**

1. Relative to the BaU control schools, what impact does the STeLLA program, as delivered using the project's scaling strategy, have on teacher knowledge and instructional practice?
2. Relative to the BaU control schools, what impact does the STeLLA program, as delivered using the project's scaling strategy, have on student science achievement?

**Moderating/differential impact:**

3. To what extent does the impact of STeLLA on teacher outcomes differ by teacher or school characteristics?
4. To what extent does the impact of STeLLA on student outcomes differ by student, teacher/classroom, or school characteristics?

**Mediating effects:**

5. To what extent is the impact of STeLLA on student science achievement mediated by teacher content knowledge, pedagogical content knowledge, and classroom practice?

**Implementation of STeLLA:**

6. To what extent is the project's strategy to scale implemented as intended (i.e., the blended version as delivered by local PD leaders)
7. What contextual factors enhance or impede the implementation of STeLLA at scale?

**Cost effectiveness:**

8. What is the cost effectiveness of STeLLA when scaled to an online format and to delivery from local PD leaders?

### E.1. Impact Evaluation Designed to Meet What Works Clearinghouse (WWC) Standards Without Reservations

The impact evaluation of STeLLA will use a cluster RCT that is designed to minimize attrition and be free of confounds. The evaluation is expected to provide evidence that will meet WWC group design standards without reservations. BSCS and its local partners will recruit 50 elementary schools, hoping to retain at least 40 in the analytic sample to detect effects as small as 0.25SD (see Power Analysis in the Technical Appendix). Assuming three teachers per school and 20 students per teacher, this will result in a sample for the impact study that includes 150 teachers and 3000 students (not including teachers and students in the pilot study). This analytic sample will be drawn from the list of partner districts in the Technical Appendix. AIR will collect baseline data on those schools and randomize them into treatment and control conditions. School-level randomization is appropriate for this intervention because STeLLA professional development is designed to facilitate within-building collaboration among grade-alike teachers through its study group structure during the academic year. To assess the impact of STeLLA on student achievement (RQ 2), we will estimate a hierarchical linear model (HLM) to account for the clustering of students (level 1) within schools (level 2) (Raudenbush & Bryk, 2002).

BSCS and AIR will minimize potential threats to the success of the evaluation. To ensure that the study does not include student joiners, AIR will establish the baseline sample by collecting rosters (as soon as available) of rising fourth graders placed into each study teacher's 2021–22 class. BSCS Science Learning and AIR will use recruitment and data collection strategies to minimize both overall and differential attrition (i.e., clear communication prior to randomization, financial incentives for data collection activities, and use of local liaisons; Roschelle et al., 2014). All treatment teachers will receive free PD as well as a $[redacted] stipend for participation in PD and data collection; all travel expenses will be paid. Control teachers will each receive $[redacted] for participation in data collection activities.

## E.2. Generation of Guidance About Effective Strategies Suitable for Replication

The proposed evaluation will generate guidance about effective strategies for implementing and scaling STeLLA in diverse settings by (a) including a large school sample representing diverse settings; (b) deliberately assessing whether the impact of STeLLA differs for different types of students, teachers, classrooms, and schools; (c) collecting and analyzing multiple sources of rich data on program implementation and scaling considerations, and (d) including a cost analysis to provide valuable information about the cost effectiveness of the program.

**Diverse settings.** The commitment of 19 partner districts that include a large number of high-need schools in diverse settings will allow the evaluation to generate valuable guidance for future replications of STeLLA in a variety of settings. (See the Technical Appendix for letters of commitment and for the demographic characteristics of the partner districts.)

**Differential impact analyses.** The evaluation will include differential impact analyses (RQs 3 and 4) to assess the extent to which STeLLA's impact is moderated by the characteristics of students, teachers/classrooms, and schools. Results from these exploratory analyses will be crucial in guiding future efforts to scale STeLLA, as they may identify settings and populations for which the program is particularly effective or not well suited. Potential moderators that will be tested include those at three levels:

1. Student level: race/ethnicity, gender, eligibility for free or reduced-price lunch (FRPL), English learner status, special education status, and prior achievement.
2. Teacher/classroom level: teaching experience, education level, class size, and classroom average prior achievement.

3. School level: school size, prior achievement levels, and demographic composition (e.g., percentage of minority students and students from low-income families, rural/ town/ suburban/ urban designation).

**Multiple sources of implementation data.** The evaluation will be informed by multiple data sources. These include those collected by BSCS and AIR for formative and summative purposes. Implementation measures include observations and/or video recordings of PD sessions and classroom instruction, a survey of teacher characteristics, curriculum implementation, instructional practices, teaching logs, and selected interviews with teachers and professional development providers; each with at least a partial focus on documenting implementation requirements and challenges of scaling to an online format and to delivery from local PD leaders.

The teacher survey collected from all treatment and control teachers three times per year and will provide data about teachers' background, teaching practices, and PD experiences (i.e., content and use of video analysis). Additional questions for treatment teachers will measure their use of STeLLA materials and strategies. The teacher survey—along with data on teacher participation in PD, observations of a sample of PD events, and artifacts from PD events—will enable us to measure the implementation of key project components (RQ 6), identify factors that facilitate or hinder the implementation of STeLLA at scale (RQ 7), and assess the achieved "service contrast" with the comparison group. Interviews with a sample of teachers in each condition will provide more in-depth data on teachers' experiences and school contexts.

**Cost effectiveness.** To provide information about whether STeLLA is a cost-effective investment and identify ways to make it more cost effective, we will conduct a cost analysis using the resource cost model (RCM), which has been used extensively by AIR (see Levin 1983; Levin and McEwan 2001). Focusing on both personnel and non-personnel resources used in

STeLLA, we will populate the RCM using the *CostOut* tool and generate cost-effectiveness estimates based on the cost estimates and results from the impact analyses.

### E.3. Providing Valid and Reliable Data on Relevant Outcomes

All data collection activities will be conducted identically across both study conditions to enable unbiased estimates of the contrasts between STeLLA and the BaU control group. To provide a valid, reliable measure of student science achievement, AIR will administer science assessments at baseline and posttest. State science assessments will provide data on policy-relevant student outcomes, as both Kentucky and Tennessee administer 4$^{th}$ and 5$^{th}$ grade science tests. Teacher content assessments, a video analysis task focused on PCK, and instructional practice videos will be administered at baseline and used as teacher outcomes and in mediation analyses.

**Study-administered science assessments.** The study-administered science assessments will be those used in the recent impact study of STeLLA (Taylor et al., 2017). These two assessments include 24 multiple choice items and cover four different content areas (Earth's changing surface, food webs, the sun's effect on climate, and the water cycle), with Rasch person reliabilities for students ranging from .74 to .77 and associated Cronbach's alphas ranging from .77 to .81 (Taylor et al., 2017). The baseline and postintervention versions of each assessment are identical. In prior studies, BSCS requested a waiver of parental consent for this student data collection because the assessments are deemed normal education practice and pose minimal risk.

**State-administered science assessments.** AIR will collect student-level data from state-administered science assessments; WWC standards indicate that state assessments are assumed valid and reliable (IES, 2017). Students' data from the prior spring and study-administered science assessments will provide baseline data. Using z-scores, tests will be equated across grades and states to facilitate use of the entire sample in the analyses.

**Teacher content and pedagogical content knowledge (PCK) assessments.** Teacher content knowledge will be assessed pre- and post-intervention (pre-summer institute, post spring semester) with the assessments used in the recent impact study of STeLLA (Taylor et al., 2017). These four assessments each have 24 multiple choice items and address the same topics as the student assessments. The Rasch person reliabilities ranged from .68 to .80, respectively, with associated Cronbach's alphas ranging from .73 to .82 (Taylor, et al., 2017). A video-based lesson analysis task will be used to measure teachers' PCK. Teachers will watch video clips of science teaching and respond to a prompt asking them about the science content, the teaching, and the students. Responses will be coded for attention to specific STeLLA strategies, as well as teachers' understanding of more general aspects of PCK. In the previous STeLLA study, the average inter-rater reliability was 93% based on double-coding of 10% of the data.

**Videos of instructional practice.** AIR researchers will collect videos of classroom instruction at three points in time for impact study participants (cohorts 2 and 3)—baseline, fall, and spring—for all teachers in the treatment and control schools. Parents of all students in study teachers' classrooms will be asked to consent to being videotaped for the study; the professional videographers will not film any students who do not give consent. All told, researchers will code three videos per teacher (fall, winter, and spring) with a coding scheme focused on enactment of STeLLA strategies (intervention fidelity) and three-dimensional science instruction (treatment–control contrast in the quality of science instruction). Scores for spring videos will serve as teacher outcomes, while scores from all videos collected from treatment teachers will be aggregated for a broader, school-year estimate of fidelity to STeLLA strategies. Post-intervention video scores also will be used in mediation analysis.

## E.4. Articulation of Program Components, Mediators, Outcomes, and Thresholds

The intended relationship among program components, mediators, and outcomes is illustrated by the STeLLA theory of influence and conceptual framework (Exhibits 2 and 3). We hypothesize that three teacher outcomes mediate the effect of STeLLA on student outcomes. If the study detects a statistically significant impact of the STeLLA model on student outcomes, we will estimate such mediation effects. To do so, we will adopt the mediation conceptualization and analytic framework of Pituch, Murphy, and Tate (2010), which will decompose the total effect into a direct effect and three indirect effects (through the mediators) and test the statistical significance of each. Key program components and the thresholds for acceptable implementation of these components are provided in Exhibit 9. To be implemented with fidelity, all three components of the program (and all indicators within each component) must meet the relevant thresholds. **Exhibit 9. Key Components and Thresholds for Acceptable Implementation**

| Components and Indicators | Threshold | Data Source |
|---|---|---|
| **Component 1: Duration and breadth of participation in PD** | | |
| Indicator 1.1. Summer institute | 90% of teachers attend at least nine of 10 all-day PD sessions | Attendance trackers/back-end user data (for online) |
| Indicator 1.2. Ongoing small-group sessions | 90% of teachers have two or fewer absences | Attendance trackers/back-end user data (for online) |
| **Component 2: Content of PD** | | |
| Indicator 2.1. Focus on content storyline lens and student thinking lens | No less than 60% of PD instructional time spent on STeLLA lenses and strategies | Artifact analysis, observations, teacher survey |
| Indicator 2.2. Focus on science content | No less than 30% of instructional time spent on deepening science content knowledge | Artifact analysis, observations, teacher survey |
| Indicator 2.3. Use of video analysis | Video analysis used in 90% of project meetings | Artifact analysis, teacher survey observations |
| **Component 3: Curricular resources** | | |
| Indicator 3.1. STeLLA strategy booklet, videocases, and model lessons | 100% of teachers use STeLLA materials as part of the PD | Teacher survey, observations |
| **Component 4: STeLLA-informed instruction** | | |
| Indicator 4.1 Use of STeLLA strategies | 90% of teachers are implementing STeLLA strategies as intended. | Classroom observations |

## REFERENCES


Ball, D., & Cohen, D. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession* (pp. 3-32). San Francisco, CA: Jossey-Bass.

Borko, H., Jacobs, J., Koellner, K., & Swackhamer, L. (2015) *Mathematics professional development: Improving teaching using the problem solving cycle and leadership preparation models*. New York, NY: Teachers College Press.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–41.

Community for Advancing Discovery Research in Education (CADRE). (2017). *Emerging Design Principles for Online and Blended Teacher Professional Development in K-12 STEM Education*. Waltham, MA: Education Development Center, Inc. Retrieved from http://cadrek12.org/resources/emerging-design-principles-online-and-blendedteacher-professional-development-k-12-stem.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development.* Palo Alto, CA: Learning Policy Institute.

Darling-Hammond, L., & Sykes, G. (1999). *Teaching as the Learning Profession: Handbook of Policy and Practice.* San Francisco, CA: Jossey-Bass.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38,* 181-200. doi:10.3102/0013189X08331140

Dorph, R., Goldstein, D., Lee, S., Lepori, K., Schneider, S., & Venkatesan, S. (2007). *The status of science education in the Bay Area* [Research brief]. Berkeley, CA: Lawrence Hall of Science.

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.

Elmore, R. F. (2002). *Bridging the Gap Between Standards and Achievement.* The Imperative for Professional Development in Education. Albert Shanker Institute.

Fulp, S. L. (2002). *2000 National Survey of Science and Mathematics Education: Status of elementary school science teaching*. Chapel Hill, NC: Horizon Research.

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*, 915-945.

Guskey, T. R., & Yoon, K. S. (2009). What Works in Professional Development? *Phi delta kappan*, *90*, 495-500. http://dx.doi.org/10.1177/003172170909000709

Hawley, W., & Valli, L. (1999). The Essentials of Effective Professional Development: A New Consensus. In L. Darling-Hammond, & G. Sykes (Eds.), *Teaching as the Learning Profession: Handbook of Policy and Practice* (pp. 127-150). San Francisco, CA: Jossey-Bass.

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher, 42*(9): 476-487.

IES. (2017). WWC procedures and standards handbook (version 3.0). Washington, DC: US Department of Education, Institute of Education Sciences, What Works Clearinghouse.

IES. (2019). *WWC procedures handbook* (version 4.0). Washington, DC: US Department of Education, Institute of Education Sciences, What Works Clearinghouse.

Kennedy, M. M. (2016). How Does Professional Development Improve Teaching? *Review of Educational Research*, 86(4), 945–980. https://doi.org/10.3102/0034654315626800

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence. *Review of Educational Research* [Internet], *88*(4): 547-588.

Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life.* Cambridge, England: Cambridge University Press.

Levin, H. M. (1983). *Cost-effectiveness: A primer* (Vol. 4). Sage Publications, Inc.

Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: methods and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Little, J. W. (2011). *Research on professional development: Implications for equity and policy* [Briefing paper]. Washington, DC: U.S. DOE Equity and Excellence Commission.

Little, R. J., & Rubin, D. B. (2002). Statistical analysis with missing data. New York, NY: John Wiley & Sons.

National Assessment of Educational Progress (NAEP). (2015). *2015 Science Assessment*. Retrieved from https://www.nationsreportcard.gov/science_2015/#?grade=4

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: The National Academies Press.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New

K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.

Osborne, J., (2014). Teaching Scientific Practices: Meeting the Challenge of Change. *Journal of Science Teacher Education, 25*, 177-196.

Pituch, K. A., Murphy, D. L., & Tate, R. L. (2009). Three-level models for indirect effects in school-and class-randomized experiments in education. *The Journal of Experimental Education, 78*(1), 60-95.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological methods, 15*(3), 209.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey methodology, 27*(1), 85-96.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Reiser, B. J. (2013). *What professional development strategies are needed for successful implementation of the Next Generation Science Standards.* Invitational Research Symposium on Science Assessment: K-12 Center at ETS.

Roschelle, J., Feng, M., Gallagher, H., Murphy, R., Harris, C., Kamdar, D., & Trinidad, G. (2014). *Recruiting participants for large-scale random assignment experiments in school settings.* Menlo Park, CA: SRI.

Roth, K. J., Bintz, J., Wickler, N. I. Z., Hvidsten, C., Taylor, J., Beardsley, P. M., Caine, A., & Wilson, C. D. (2017). Design principles for effective video-based professional development. *International Journal of STEM Education. 4*:31. https://doi.org/10.1186/s40594-017-0091-2

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. Z. (2011), Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching*, *48*: 117–148

Roth, K. J., Wilson, C., & Taylor, J. (2014, April 5). *Improving teacher and student science learning through videocase-based lesson analysis*. Presented at the annual meeting of the American Educational Research Association. Philadelphia, PA.

Roth, K. J., Wilson, C. D., Taylor, J. A., Stuhlsatz, M. A. M., & Hvidsten, C. (2018). Comparing the Effects of Analysis-of-Practice and Content-Based Professional Development on Teacher and Student Outcomes in Science. *American Educational Research Journal*. https://doi.org/10.3102/0002831218814759

Shulman, L. (1987) Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, *57*(1), pp. 1-23.

Spybrook, J., Bloom, H., Congdon, R., Hill, C., Martinez, A., Raudenbush, S., & TO, A. (2011). *Optimal design plus empirical evidence: Documentation for the "Optimal Design" software*. William T. Grant Foundation.

Spybrook, J., Westine, C. D., & Taylor, J. A. (2016). Design parameters for impact research in science education: A multistate analysis. *AERA Open, 2*(1), 2332858415625975.

Taylor, J. A., Roth, K., Wilson, C. D., Stuhlsatz, M. A. M., & Tipton, E. (2017). The effect of an analysis-of-practice, videocase-based, teacher professional development program on elementary students' science achievement. *Journal of Research on Educational Effectiveness, 10:*2, 241-271.

Westine, C. D., Spybrook, J., & Taylor, J. A. (2013). An empirical investigation of variance design parameters for planning cluster-randomized trials of science achievement. *Evaluation Review, 37*(6), 490-519.

Wilson, C. D., Stuhlsatz, M., Hvidsten, C., & Stennett, B. (2017, January). *Videocases for Science Teaching analysis Plus (ViSTA Plus): Initial Findings from a 3-year program preparing elementary teachers to teach science*. Presented at the International Conference of the Association for Science Teacher Education, Des Moines, IA.

Wilson, C. D., Taylor, J. A., Roth, K., Stuhlsatz, M. A. M., & Hvidsten, C. (2016). *The Effect of an Analysis-of-Practice, Videocase-Based, Teacher Professional Development Program on Teacher and Student Outcomes.* Paper presented at the Annual Conference of the National Association for Research on Science Teaching, Baltimore, MD.

Wilson, S. M. (2011). *Effective STEM teacher preparation, induction, and professional development*. Paper presented at the NRC Workshop on Highly Successful STEM Schools or Programs. Available from http://www7.nationalacademies.org/bose/Successful_STEM_Schools_Homepage.html

Wilson, S. M. (2013). Professional development for science teachers. *Science, 340*, 310-313.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement.* Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.