NARRATIVE

Making Sense of Science and Literacy:

Improving Teacher Effectiveness and Building Capacity

Table of Contents

Priorities.....1
Improving the Effectiveness of Teachers (Absolute Priority 1)
Improving Cost-Effectiveness and Productivity (Competitive Priority 1)
Enabling the Broad Adoption of Effective Practices (Competitive Priority 2)

Significance of the Project 4
National Need for Improved Student Achievement
Addressing National Need
Estimated Impact and Scale
National Expansion

Quality of the Project Design.....10
Clear Goals and Strategies
Addressing Barriers to Scale

Quality of the Management Plan.....20
Timeline and Measurable Milestones
Qualifications of Key Personnel
Multi-year Operating and Financial Model

Quality of the Project Evaluation.....25
Study Questions and Methods
Sampling, Power Analysis, and Random Assignment
Measures and Data Collection
Analysis Plan

PRIORITIES

Through a validation grant award, WestEd, a nonprofit educational organization, in partnership with Teaching Channel (Tch), an award-winning online learning environment, will address an unmet national need—improving teacher ability to implement the rigorous Next Generation Science Standards (NGSS) and meet the disciplinary reading, writing, and discourse demands of the Common Core State Standards in English Language Arts (CCSS-ELA) in ways that substantially improve the achievement of high-needs elementary students, setting them up for success in middle school and onward toward career and college readiness.

Committed LEA partners include Milwaukee Public Schools (Wisconsin) and San Joaquin County Office of Education (California) with a consortium of small and rural districts in the central farming valley. These partners meet the following criteria: (a) serve large populations of high-needs students (e.g., those at risk of educational failure, English learners, high-poverty communities), (b) are committed to improving science and literacy education by collaborating in a regional hub, and (c) confirm their teachers' needs match the intervention (Appendix G). These same criteria will apply when regions expand and/or additional districts are recruited.

A top-notch research team comprised of Empirical Education Inc., under the direction of Dr. Denis Newman, Heller Research Associates, under the direction of Dr. Joan I. Heller, and a Technical Working Group of advisors, will carry out confirmatory and exploratory studies to validate the impact of this model in these two diverse settings, evaluate the supports and obstacles to implementation, provide formative feedback to WestEd and Teaching Channel to strengthen the program and teacher resources, and evaluate the cost-effectiveness of the model.

To address **Absolute Priority 1: Improving the Effectiveness of Teachers,** WestEd will train regional teams—composed of university partners, cadres of teacher leaders, and district literacy and science specialists—to lead Making Sense of Science and Literacy (MSSL,

formerly Making Sense of SCIENCE) professional development (PD) that has been shown to strengthen teachers' knowledge and skill, and transform their classroom practices, in ways that support implementation of NGSS and CCSS. MSSL courses are carefully sequenced, yet modular by design, making them usable in a variety of contexts. Teachers who participate in these 40-hour courses learn effective practices for leading hands-on science investigations, supporting evidence-based discussions, and developing students' reading and writing skills, along with the habits of mind necessary for sense-making and scientific reasoning. Each course includes four main components:

(1) *Science Investigations,* where teachers strengthen their content knowledge, inquiry skills, and ability to identify and correct common misconceptions;

(2) *Literacy Investigations,* where teachers learn to implement classroom practices that strengthen students' abilities to write, read, and talk in science-specific ways;

(3) *Teaching Investigations,* where teachers examine student thinking and explore instructional practices to help move students toward more accurate understandings;

(4) *Classroom Connections,* where teachers reflect upon key concepts and consider how these pertain to their own work with students. (See Appendix J for sample materials.)

To extend the reach of highly effective teachers (subpart 2), we will build an open-source, digital library of student work exemplars and videos of effective practices available through Teaching Channel's innovative, technology-enabled learning environments, and train teacher leaders to facilitate online professional learning communities (PLCs) using these resources.

The project will also address **Competitive Preference Priority 1: Improving Cost-Effectiveness and Productivity** by evaluating the gain in productivity achieved through the MSSL approach to dissemination that draws on the knowledge and skill of a district's most

effective teachers and utilizes an efficient train-the-trainer model in conjunction with regional hubs that build local capacities. To determine the measurable return on investment and establish the cost-effectiveness of MSSL, we will conduct an exploratory study that employs the widely-used ingredients method (Levin & McEwan, 2001) to tabulate program inputs and costs in alternative programs. To investigate strategies that could further reduce costs and facilitate broader scale-up, we will inspect key elements of program delivery with an emphasis on exploring the possibility of achieving cost savings via online program delivery.

Competitive Preference Priority 2: Enabling the Broad Adoption of Effective Practices will be tackled in a systemic and sustainable way by (a) establishing regional hubs comprised of district leaders, university faculty, and partners from business and philanthropy who work together to set goals, align resources, involve key stakeholders, and ensure ongoing support to teachers through access to high-quality professional learning opportunities; (b) developing digital tools and resources (e.g., classroom artifacts, videos showcasing effective practices, protocols for peer-to-peer feedback); and (c) refining the model of dissemination by implementing MSSL in a variety of settings to identify and hone key elements essential to broad adoption, national scale-up, and sustainability. Anticipated outcomes include:

- **At least 60 highly effective teachers trained** to facilitate Making Sense of Science and Literacy PD across large urban districts and nearby rural schools in CA and WI.
- **More than 800 teachers** with stronger content knowledge and more effective pedagogy who **improve the science and literacy achievement of at least 20,000 elementary students.**
- **Important contributions to research and education** with respect to (a) the impact of MSSL scale-up on teacher effectiveness and student achievement, revealed by an RCT, (b) successful strategies for extending the reach of highly effective teachers, and (c) the utility of regional hubs in fostering scale-up, with insight into obstacles and supports.

SIGNIFICANCE OF THE PROJECT

National Need for Improved Student Achievement. With job growth taking place in occupations such as research, health care, and drug discovery, scientific literacy is increasingly essential (Kasper, 2006; Bureau of Labor Statistics, 2012). To be successful in these careers, students must engage in deep critical thinking, inquiry, problem solving, and teamwork. They must also be able to read, write, and speak with fluency in science. This includes being able to comprehend informational texts, generate and support explanations based on evidence and reasoning, and translate between information presented in a text, model, graph, or table.

Unfortunately, the majority of our youth fall far short on these essential skills. Results from the 2013 National Assessment of Education Progress (NAEP) show only 1 in 4 of California's 4th grade students are prepared, with 27% scoring at or above proficient in reading, 23% in writing, and 22% in science (NCES, 2014). On the same test, Wisconsin's 4th graders fared only slightly better (Ibid). Students with the lowest achievement are far more likely to be from high-needs communities, students of color, and/or non-native English speakers (NCES, 2012).

New standards, such as NGSS and CCSS raise the bar for what it means to be proficient in science and literacy. At the same time, these rigorous standards are viewed as an opportunity to close achievement gaps, especially for Hispanic and Black students in the U.S. who perform especially poorly on international assessments (Baldi, Jin, Skemer, Green, & Herget, 2007). Yet, many teachers are woefully unprepared to address these standards (Banilower, Smith, Weiss, Malzahn, Campbell, & Weiss, 2013). Schools with the greatest numbers of high-needs students are at the greatest disadvantage. They hire teachers with weaker qualifications in terms of experience, certification, and post-baccalaureate coursework (Presley, White, & Gong, 2005).

In numerous studies, teachers have been shown to be the single most important variable impacting student achievement (Duschl, Schweingruber, & Shouse, 2007; Hill, Rowan, & Ball,

2005). More effective teaching has been linked to mentorship from experienced teachers, on-the-job experience, and a bachelor's degree in the subject (Leana, 2011). Thomas Kane, a Harvard economist, predicted if we train teachers in the practices of highly effective teachers, we could raise the average classroom achievement to that of the top quarter (Green, 2010).

Unfortunately, few programs focus on improving teachers' pedagogical knowledge for teaching science. Most solely focus on content or on classroom management (Sztajn, Marrongelle, Smith, & Melton, 2012). High-needs school districts often lack the resources to support coherent, sustained, and effective PD (Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009; Desimone, 2009). Yet, teachers who have access to quality PD feel more prepared and professionally connected, are more effective, and are likely to remain in the classroom longer (Hanushek, Kain, & Rivkin, 2004; Ingersoll & May, 2010; Yost, 2006).

Addressing National Needs. MSSL is a research-based intervention poised to tackle this critical need. Developed at WestEd, with support from NSF, the Institute of Education Sciences (IES), the Stone Family Foundation, and the Stuart Foundation, MSSL has been informed by more than a decade of research and development, involving thousands of teachers, dozens of scientists and science educators, and literacy specialists from the Strategic Literacy Initiative and National Writing Project (both i3-supported initiatives). In 2012, Change the Equation, a business-supported, White House initiative, designated MSSL a proven program worthy of inclusion in its selective STEMworks database.

The fundamental goal of MSSL is to improve students' understanding of science and literacy skills, especially for English learners and low-performing students, by strengthening the content knowledge and instructional skill of their teachers. The theory of action is based on the premise that when professional development is situated in an environment of collaborative inquiry, this leads to increases in teachers' content and pedagogical content knowledge, which

results in changes in classroom practices (e.g., increased accuracy of science, a focus on conceptual understanding, and opportunities for students to talk/write to learn science). These classroom changes produce improvements in student achievement, as shown in Figure 1.

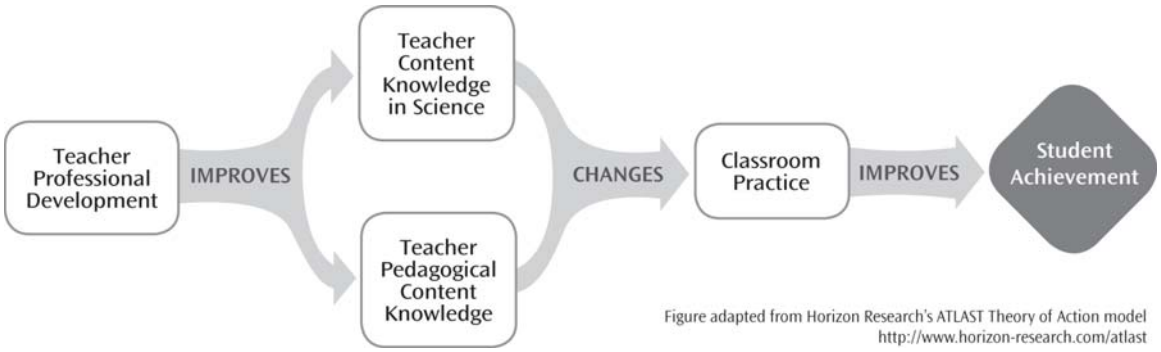

Figure 1. *Theory of Action linking MSSL professional development and student achievement.*

Estimated Impact and Scale. Important research has gone hand-in-hand with development of the MSSL model and materials, with a series of increasingly rigorous quasi-experimental and experimental studies. In each study, individual MSSL courses have been shown to improve student achievement in measurable ways—for example, statistically significant differences were found favoring treatment teachers and students on measures of science content knowledge, with non-native English speakers and low-performing students making the greatest gains (Heller, Daehler, & Shinohara, 2003; Heller et al., 2012a, 2012b). This impact has been found for different MSSL courses, across multiple states, for native English speakers as well as limited-English-proficient learners, and with students from a range of socioeconomic backgrounds. Collectively, these data provide strong evidence of the internal validity of the model. They also offer strong evidence of external validity with respect to the model's impact on increasing students' achievement and showing promise for **reducing the achievement gap**. However, previous research has not yet tested the program's impact when scaled up as proposed here, nor in the context of supporting district-wide implementation of either the NGSS or CCSS-ELA, both of which require new ways of teaching and a redefinition of effective practices.

One of the most rigorous tests showing the positive impacts of the MSSL model was a randomized controlled trial (RCT), conducted over a two-year period (2007–2009), in eight sites across six states in the U.S., involving 49 districts, more than 260 elementary teachers and nearly 7,000 students, largely from underserved populations. This study, carried out by an external team of researchers with support from NSF found **treatment effects were statistically significant for all major demographic groups**, with non-native English speakers gaining more than native speakers, and low-performing students making the biggest gains, as shown in Figure 2. Treatment teachers made exceptional gains in their science knowledge (ES = 1.8, $p < .001$). Students of these teachers outperformed students of control teachers by more than 40 percent (ES = .36, $p < .001$). In 2008–2009, a second RCT was conducted at six sites with over 130 teachers and approximately 6,000 students, and comparable benefits to teachers and their students were found. (See Appendix D for details.)

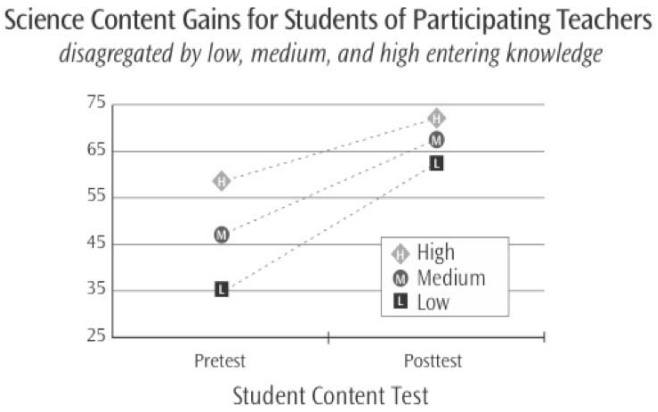

Figure 2. *Greatest gains by low achievers.*

Based on prior results, we estimate this i3 effort will positively impact student achievement with an effect size of at least .36 over the course of the two-year intervention. The intervention is also likely to increase teacher retention, by enabling them to be better prepared and connected to a vibrant local and online community, thus benefiting a greater number of students each year the teacher remains in the classroom. More specifically, this i3 project will significantly impact students in high-needs urban and rural regions, including:

- **San Joaquin (CA)**—a migrant farming area with many rural schools facing challenges of isolation and lack of resources (e.g., limited access to technology, professional development,

and informal science). The surrounding counties are home to many limited-English-proficient learners and a staggering number of families living below the poverty line. With soaring unemployment the largest city in the region, Stockton, has struggled financially and become the largest U.S. city to file for bankruptcy until surpassed by Detroit in 2012.

- **Milwaukee (WI)**—home to Milwaukee Public Schools (MPS), a large, diverse, urban district plagued by economic duress, a rapidly growing population of English learners, and a host of challenges that accompany poverty. In 2013, a mere 15% of MPS students were proficient in reading and only 19% were proficient in math according to the state test. That same year, nearly 30% of MPS students did not graduate from high school in four years.

National Expansion. Over the past decade, demand for MSSL has grown and the program has expanded into more than 20 states, reaching thousands of teachers, providing tens of thousands of hours of PD, and impacting hundreds of thousands of students. Since 2004, several NSF and IES grants have supported development and national field-testing of many new MSSL courses covering topics in earth, life, and physical science for K–8 teachers (e.g., energy, organisms), the first of which have been co-published and widely disseminated by the National Science Teachers Association (NSTA) with more slated for release in 2015. Recent, prestigious awards include:

- Selected as provider-of-choice for three consecutive years to train 40 educators from the Texas Regional Collaboratives, impacting more than 2000 teachers and 10,000 students.
- Presented WestEd's "Paul D. Hood Award for Distinguished Contribution to the Field" in 2013 for outstanding work in research, development, and service.
- Chosen as a gold-standard PD model for the Smithsonian's 2010 i3 validation grant, to develop 31 mini MSSL courses, impacting 1,300 teachers across NC, NM, and TX.

The MSSL model embodies many of the characteristics of innovations that have successfully scaled up, including a high degree of *focus, speed* in achieving short-term results that support

long-term objectives, *coherence* of elements working together to meet overall goals, and *comprehensiveness* in its goals for long-term impact. In addition, MSSL is replicable in a variety of settings, implemented at the district/school level, cost-effective, not human capital-intensive (employing a train-the-trainer model), and achieves significant results after only one year of implementation. In a recent national randomized study, the MSSL intervention was shown to be equally effective when led by trained regional facilitators, as when led by WestEd developers, with comparable teacher and student gains (Heller et al., 2012b). Over the past few years the project has invested significant resources into scaling—hiring new staff, training new facilitators, and assisting regions in building their local capacities and aligning resources.

In terms of scale, a single MSSL Facilitation Academy can result in 20 trained teacher educators (e.g., content partners, teacher leaders). When these teacher educators work in pairs with 24 teachers, then these 24 teachers with their enhanced knowledge and skill, can return to their schools impacting an average of 25 elementary students each year they remain in the classroom. In this manner, a **single Facilitation Academy can indirectly affect the science learning of 3,600 students in one year.** (See Figure 3.)

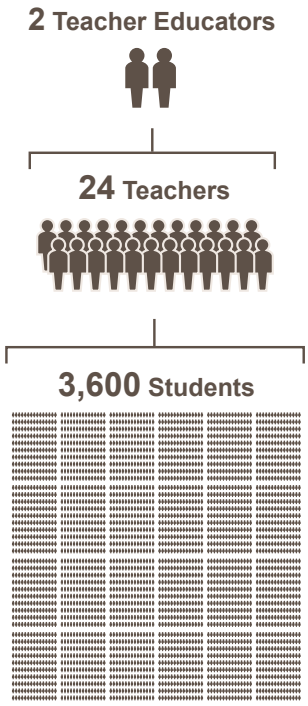

Figure 3. By training two teacher educators who lead a single course, up to 3,600 students can be reached in a single year. When more courses are led, the impact multiplies.

If these more effective teachers remain in the classroom the following year, another 3,600 students will benefit. Following this logic, it is reasonable to assume a single WestEd instructor could lead four or more Facilitation Academies in one year, such that nearly 16,000 elementary students (or 100,000 middle school students) would benefit.

QUALITY OF THE PROJECT DESIGN

Clear goals and strategies. With support from an i3 grant, MSSL will accomplish the following goals by employing targeted strategies, identifying potential risks, and mitigating those risks by modifying plans and refining approaches. Our strategies for addressing Competitive Priorities 1 and 2 are also consistent with preparing to expand MSSL statewide, and eventually nationally.

***Goal 1:* Transform teaching and learning in ways that measurably increase elementary students' science and literacy achievement, prepare students to meet rigorous new standards, and close achievement gaps.**

To achieve this goal, the MSSL approach is shown in the logic model in Figure 4.

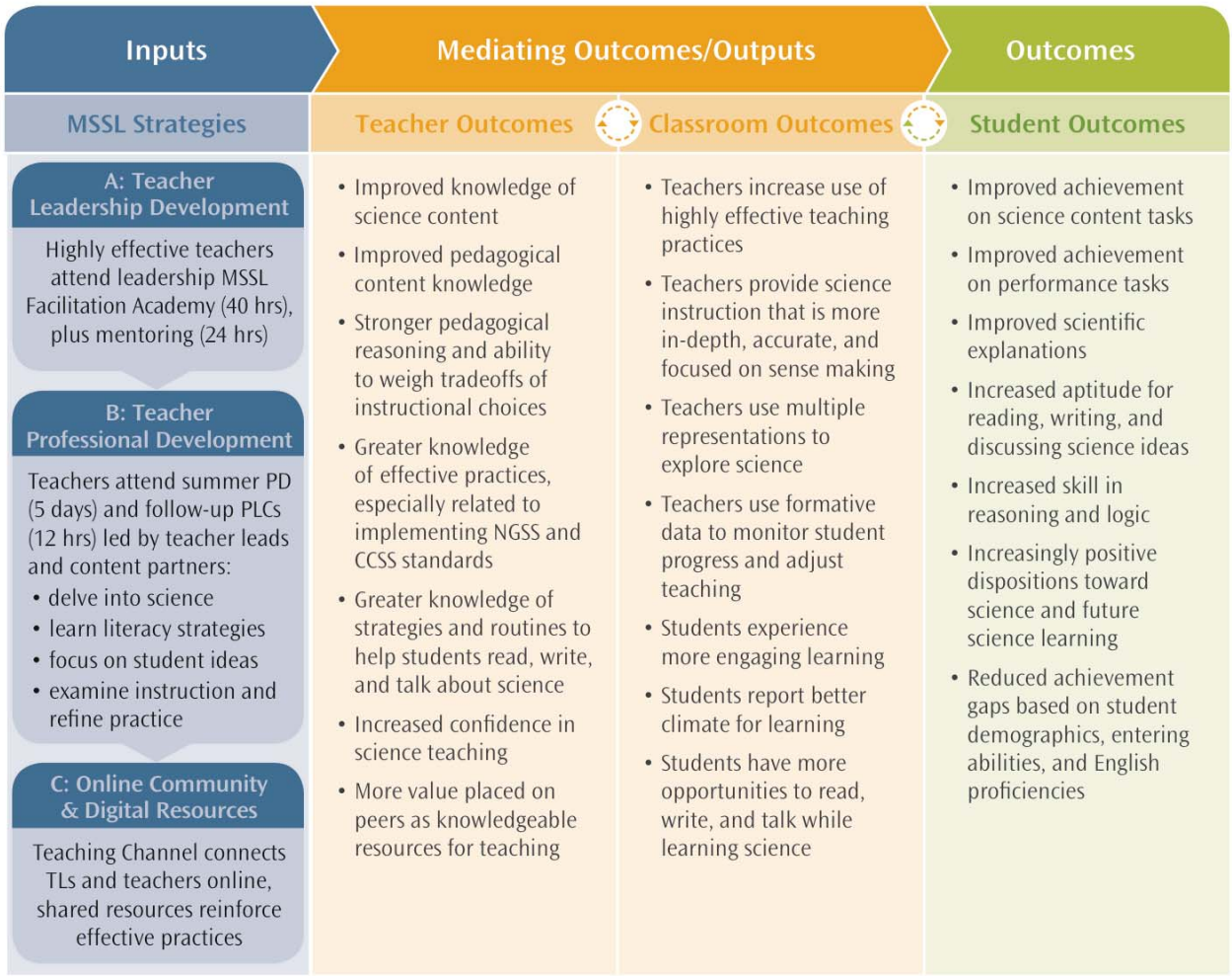

Figure 4. *Making Sense of Science and Literacy logic model addressing Goal 1.*

Strategies include providing targeted leadership development to a select group of teachers who then co-lead the MSSL intervention and provide support to their peers in onsite PLCs and

online through Teaching Channel's innovative Learning Labs. These inputs strengthen the effectiveness of hundreds of other teachers, resulting in a feedback loop with better prepared teachers contributing to improved learning experiences and better classrooms contributing to greater teacher confidence, skill, and effectiveness. As a result these mediating outcomes impact students in significant ways that feedback to improve classrooms.

Strategy A: *Train cadres of teacher leaders to provide peer-to-peer and online support.* As a mechanism for supporting 800 elementary teachers, we will begin by identifying and training a cadre of 60 teacher leaders (TLs). The research base supports this strategy as teacher leadership development yields a high return on investment. For example, Candal and Klemmer (2013) write, "We now understand, better than ever before, that . . . [effective] teachers can have an even greater impact on the success of a school when administrators thoughtfully deploy them by positioning them to lead and extend the capabilities of their colleagues." In addition, *The Model Teacher Leader Standards* describe how teacher leaders promote a culture of shared accountability "for school outcomes that maximize teacher effectiveness, promote collaboration, and drive continuous improvement in instruction to improve student learning," (Teacher Leadership Exploratory Consortium, 2010). Also, lasting cultural shifts in schools require the involvement of teacher leaders (Beachum & Dentith, 2004).

We know quality teaching happens everywhere, in pockets. To identify 1–2 effective teacher leaders (TLs) per school, the project will conduct an interview of the principals following a protocol that will be developed and administered prior to randomization in each of the schools. The school districts where the independent evaluation is to be undertaken have very different approaches to teacher evaluation and therefore provide different resources for use in identifying highly effective teachers. While Milwaukee has a formal framework involving classroom observations (based on the Danielson Framework for Teaching) and other elements

such as student growth calculations, the CA districts typically have a less formalized approach. Districts' established measures of effectiveness will provide a basis for TL selection, but will be supplemented by principal interviews using a protocol that will focus on characteristics expected to be associated with effectiveness, not just as a teacher, but also as a mentor to peers (e.g., empathetic to the challenges of new teachers, exemplary resources to share, experience observing and coaching others). The project will also consider qualifications such as National Board Certification, degrees in science, and three or more years teaching experience.

In the first year of this project, TLs will attend a five-day summer institute, receive 3 days of instructional coaching and mentoring in leadership development, and participate in 12 hours of peer-to-peer PLCs. In spring of year two, these TLs will attend an MSSL Facilitation Academy along with local content partners (e.g., museum staff, university educators), in which they will learn to lead the PD, including the MSSL principles of facilitation (e.g., making thinking visible, not stopping at one, exploring ideas with words, actions, images, and symbols).

During the summers (2016 and 2017), teams of content partners and TLs will jointly facilitate MSSL courses for teachers. During the year, the TLs will provide site-based support by leading PLCs with their peers. This PLC component is key, as a two-year study conducted by the National Commission on Teaching and America's future offers compelling evidence that "teaching is more effective and student achievement increases when teachers join forces to develop strong professional learning communities in their schools" (Britton & Fulton, 2011).

As an additional means of further extending the reach of these highly effective teachers, beginning in the second year and throughout the duration of the project, the cadre of TLs will be invited and supported in collecting artifacts of effective practice (e.g., samples of student work, teaching cases of practice, videos of instruction) that will be made available to teachers across the country via the Teaching Channel's open-source digital library. Later, a subset of these TLs

will be invited, trained, and compensated to facilitate online PLCs, as a mechanism for sharing expertise with geographically remote teachers and reducing costs of scaling the intervention.

Strategy B: *Scale up a proven professional development program.* For this i3 project, implementation will take place as a two-year science and literacy initiative, with school-wide involvement of upper elementary teachers. Starting in summer 2016, 300 teachers from the 30 randomly selected treatment schools will attend the MSSL *Matter & Energy* course, followed in summer 2017 with the *Earth Systems* course. In the fall, a one-day principal's meeting will be held to give administrators a first-hand understanding of the MSSL intervention so they are well informed and committed to supporting their teachers' efforts to transform teaching.

During the academic year, teachers will meet in PLCs for six 2-hour sessions. Using a structured protocol these study groups will examine students' work from their own classrooms, learn best-practices in formative assessment, and reflect on, evaluate, and refine their instruction. This approach incorporates characteristics of effective learning communities, including shared values and goals, leadership support, use of student data/work, trust and collective responsibility, and good facilitation (Britton & Fulton, 2011).

Strategy C: *Utilize Teaching Channel's innovative online learning environment to support teachers in sharing and examining classroom artifacts and effective practices.* Professionals build their expertise and skill when they are able to share and discuss their art with each other. Teaching Channel (Tch) is a non-profit organization that provides teachers the tools they need to do just that via an online, asynchronous collaboration platform, centered on helping teachers engage in evidence-based, job-embedded learning through the use of video. Tch has a rapidly growing community of over 525,000 registered teachers who trade ideas and share inspiration from each other in a web-based community, making it one of the top 10 websites used by teachers. Leveraging Tch's existing technologies, participating MSSL schools and teachers will

have their own Teaching Team space to organize and disseminate classroom artifacts. Access to these resources will help teachers develop a more sophisticated and robust repertoire of approaches for helping students master the NGSS and CCSS. This will allow teachers to think deeply about new practices and engage in discussion with each other, as they work to translate and/or adapt teaching practices for use in their own classrooms. In addition, TLs will have a separate space to where they can share facilitation challenges and tips, receive advise from MSSL staff, and strengthen their mentoring and leadership skills.

***Goal 2:* Enable broad adoption of effective practices regionally and build national capacity to sustain, support, and scale MSSL.**

The MSSL approach embodies a host of effective PD and scale-up practices. (See Figure 5.)

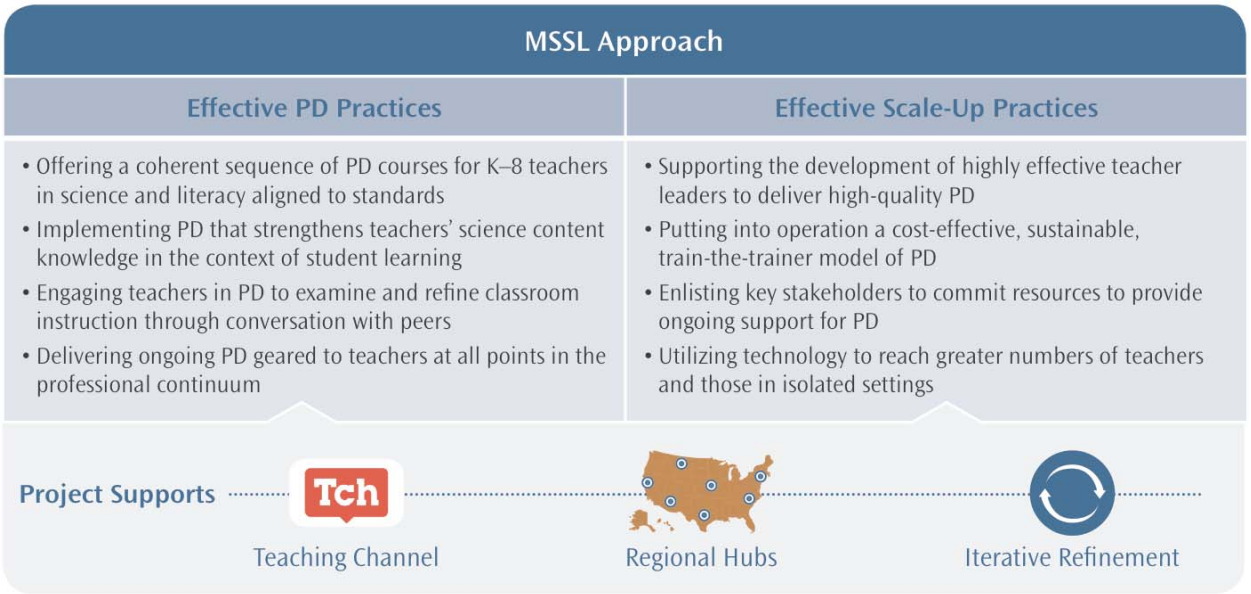

Figure 5. *Project Supports for Effective PD and Scale-up Practices Related to the MSSL Approach*

With funds from this i3 grant the project will develop new online tools and resources, implement a regional hub model previously tested (Daehler, Balster, & Heller, 2013), develop toolkits to support new hubs, and implement MSSL in a variety of settings to iteratively refine the model for scale up and sustainability.

Strategy D: *Develop online tools and a national library of digital resources to support effective implementation.* Many teachers, not just those in rural contexts, face professional isolation and limited access to high-quality learning. Technology can be an important and cost-effective piece of the solution, especially to offer resource-rich supports and critical guidance to teachers as they implement the NGSS and CCSS. We will develop a new feature within Tch's Learning Labs to support the PLC component of the PD. This will allow teachers to upload, examine, annotate, and discuss student work. The experience will be interactive and guide peer-to-peer conversations, modes of online learning shown to be most effective for teachers (USDE, 2009) and most crucial to the lasting quality of the PD (Darling-Hammond et al., 2009).

The design principles are informed by the recent report, *Transforming American Education: Learning Powered by Technology* (Institute of Education Sciences, 2010), which includes a call to "design, develop, and adopt technology-based content, resources, and online learning communities that create opportunities for educators to collaborate for more effective teaching, inspire and attract new people into the profession, and encourage our best educators to continue teaching" (p. 11). Further, the design blends online with in-person learning, as studies show this is more effective than traditional or online-only models (USDE, 2009). This new Tch feature will be developed over three years, with TLs involved in conceptualizing and testing. Later, TLs will be invited, trained, and compensated for facilitating the online Learning Labs.

In addition, we will build an open-source digital library of resources that will allow other teachers across the country to examine samples of student work and view videos of effective science and literacy practices in action in real classrooms. Participating teachers will be invited to contribute student work samples and video crews will capture exemplars of their teaching practices. Prior to publishing resources for public use, Tch will follow a process for ensuring sufficiently high quality videos and alignment with NGSS and CCSS-ELA practices.

Strategy E: *Employ a hub model to build regional capacities by involving key stakeholders.* Similar to "hub" models used in health care and social services, we will bring together key stakeholders from districts/LEAs, higher education, business, and philanthropy (see Figure 6).

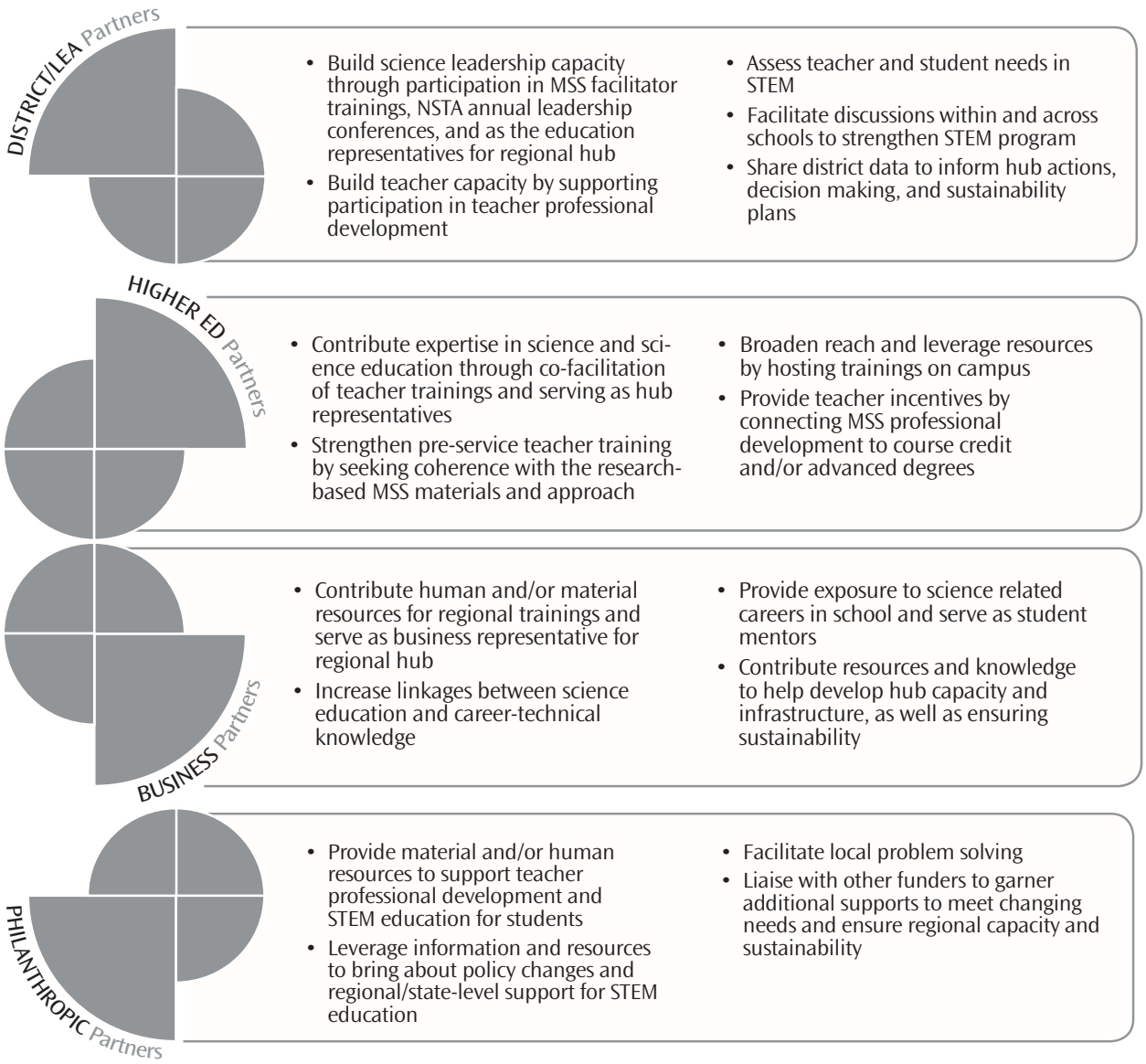

Figure 6. *The support roles of collaborative hub partners.*

Working as a collaborative, these hub partners will take on various support roles such as: conducting needs assessments, setting regional STEM education goals, aligning resources, and promoting local empowerment and autonomy. In an earlier study (Daehler, Balster, & Heller, 2013), WestEd established four regional hubs, involving more than 80 members who shared the

mission of strengthening earth science education. These hubs contributed to the development of materials for teacher learning and facilitated knowledge and resource exchange among sites.

We have already identified key hub partners, including businesses with an interest and commitment to the effort, as well as faculty at colleges and universities. Leaders in each hub will meet every other month to plan and reflect on their progress. HRA will provide formative feedback to inform planning. Hubs will also meet periodically with State Education Agency members to discuss policy barriers and best practices. In this way, states can leverage what is learned to support education broadly, affect policy, and contribute to sustainability and spread.

With i3 funds we will develop a toolkit for hubs to support capacity building and MSSL implementation. This resource will help guide future hub leaders in their outreach and sustainability of their efforts. Other targeted leadership tools will include archived webinars.

Strategy F: *Implement MSSL in a variety of settings to identify and hone key elements essential to broad adoption, national scale-up, and sustainability.* In summer 2018, phase II of MSSL implementation will begin with 300 teachers from control schools, plus 200 teachers from remote rural schools, at the grade levels (teachers 3–8) and topics of the districts' choice. During this phase, we will test the hybrid model of MSSL, such that teachers will experience the 5-day summer institutes in person and then participate in online PLCs. New regional hubs will also allow the project to build on lessons-learned in order to refine the model.

Goal 3: **Reduce the costs of achieving beneficial outcomes for teachers and their students.**

Strategy G: *Monitor costs associated with implementing an economical train-the-trainer approach, evaluate its cost-effectiveness in comparison to alternative programs, and identify elements with potential for additional cost reduction.* We will achieve this through a rigorous analysis of student outcomes and analysis of associated costs in treatment and control schools. Since a confirmatory study will establish whether MSSL substantially improves student

[redacted]

study is not designed to estimate cost-effectiveness impact of particular elements of MSSL, the data will allow for formulating hypotheses and design choices for a subsequent study.

For the purpose of detailed cost comparisons, we will survey school and district administrators in the control group about the direct costs incurred in facilitating science instruction (e.g., PD activities, other initiatives, and material costs). For MSSL, which is curriculum independent, we will amortize the costs over the average expected tenure of participating teacher (extrapolated from current teacher career data). For programs or materials associated with a particular curriculum, we will amortize the costs over the typical time between curriculum changes (based on surveys of administrators).

In addition, we will calculate the implicit costs of science instruction in both treatment and control schools by imputing the cost of instructional time (following the approach developed in Lazarev & Newman, 2011). This is necessary to account for potential variability in idiosyncratic (individual) and systematic (school-level and program-induced) propensity to allocate extra time for science instruction. We will collect the following data: number of hours of science instruction, S (based on teacher surveys); school operating annual costs (net of direct program program cost), V (from school budget data); total number of instructional student-hours per year, H; class size, N (from school administrative records). From this data, we will impute the annual "time cost" of science instruction as $Ct = (V/H) * S * N$ for each participating teacher (where V/H is implicit cost of one student-hour of instruction).

The summative evaluation of MSSL relative cost effectiveness will be based on the calculation of effectiveness-cost ratio for science instruction, E/C. The cost metric included in this calculation will be represented by the sum of direct program costs (discussed earlier) and the imputed time cost, Ct, described above. We will use student gains on NWEA tests between beginning and end-of-year administrations as a measure of program effect, E. Effectiveness-cost

ratios will be calculated for each participating teacher, averaged over treatment and control groups, respectively, and the averages will be tested for the significance of the difference.

Addressing Barriers to Scale. Over the past decade, the MSSL team has been nimble, creative, and resourceful as it has met the challenges associated with expansion (e.g., district-wide budget cuts, shifts in educational priorities, changes in policies and leadership, declines in grant funding), by creating concrete plans to address such challenges (e.g., connecting districts with grant funders, designing an affordable "train-the-trainer" model). The result is a project with a documented track record of identifying and addressing barriers to scale in order to grow and sustain the work. For this i3 project, the team has identified several potential barriers to scale, and has planned ways to use award funds to overcome them. (See Table 1.)

Table 1. *Barriers to Implementation and Approaches to Overcoming Barriers.*

Barrier	Use of Award Funds to Overcome Barrier
(1) Administrative changes. Changes in school/district leadership is common in high-need districts.	Significant resources are allocated to strengthening all teachers' effectiveness across the districts. This results in more distributed expertise and greater "staying power" as administrators change.
(2) Competing instructional demands. Many elementary school teachers instruct multiple subjects, making for competing demands on their time and focus.	To best respond to varied demands placed on teachers in the LEAs, WestEd staff will work closely with district coaches to monitor and adjust the program. WestEd will also provide NGSS/CCSS crosswalks so teachers can see how science instruction can help students develop other academic skills.
(3) Teacher turnover. Over the five-year duration of this project, some teachers may change grade levels and/or switch schools.	In multi-year, large-scale education efforts, teacher mobility is an issue. Anticipating this, WestEd will implement MSSL across grades 3–5, as teachers often shift among these upper elem grade bands. We will also foster cross-grade PLCs for info sharing.
(4) Sustaining change. It takes time to build a program that has merit and meaning over time, and beyond grant funding.	Building local infrastructure diminishes the need for ongoing WestEd support, so hubs will begin planning for sustainability in year one. To support new hubs after the grant period, we will develop a hub toolkit. The Tch portal will provide ongoing, national-wide access to exemplar teaching resources.

QUALITY OF THE MANAGEMENT PLAN

WestEd is a preeminent educational research, development, and service organization with 600 employees and 16 offices nationwide. WestEd has been a leader in moving research into practice by conducting research and development (R&D) programs, projects, and evaluations;

by providing training and technical assistance; and by working with policy makers and practitioners at state and local levels to carry out large-scale school improvement and innovative change efforts. WestEd has a strong history managing complex projects, with nearly 50 years of experience working directly with LEAs on over 2,000 successful projects nation-wide.

Timeline and Measurable Milestones. To assure the effective execution of this i3 grant, WestEd has constructed a comprehensive plan that includes measureable milestones, actions, and outcomes, annual performance targets, and the metrics to assess progress. (See Table 2.)

Table 2. *Summary of key milestones between January 2015 and December 2019.*

	Teaching & Learning	Research & Evaluation	Capacity Building
Y1	• Interview principals to recruit 60-90 teacher leaders (TLs). • Lead MSSL for TLs (phy sci), plus 6 school year PLCs (f2f). • Plan collection of resource for Tch digital library (classroom exemplars, student work). • Conceptualize online PLCs for Tch platform with TLs.	• Update eval plan, complete IRB. • Obtain district MOUs, recruit and randomize 60 schools for RCT. • Adapt and refine instruments. • Conduct principal interviews. • Obtain school-level baseline data. • Obtain teacher and TL baseline data (PASS1, survey, interviews). • Interview hubs and share feedback.	• Identify 10–15 key stakeholders per hub, including content partners (univ, museum). • Kick-off hub mtg to build awareness, identify needs, and set goals. • Quarterly hub mtgs to check progress & plan.
Y2	• Leadership develop for TLs. • Lead Facil Academy (phy sci) for TLs and content partners. • TLs/content partners lead MSSL for 300 Txt Ts (phy sci), plus school year PLCs (f2f). • Lead MSSL for TLs (Earth sci), plus 6 school year PLCs (f2f). • Collect digital library resources. • Pilot online PLCs with Tch.	• Collect TL follow-up data (surveys, interview). • Collect/analyze implement data. • Develop, pilot, and validate SQI protocol for video obs. • Interview hubs and share feedback. • Provide implementation feedback to MSSL program staff.	• Eval meet with hubs to document challenges & accomplishments, suggest modifications to increase impact, and capture lessons learned. • Hub leaders meet quarterly to adjust plans, share strategies, and build network.
Y3	• Lead Facil Academy (Earth sci) for TLs and content partners. • TLs/content partners lead MSSL for 300 Txt Ts (Earth), plus 6 school year PLC mtgs. • Train online PLC facilitators. • Collect digital library resources.	• Collect TL data (survey, interv). • Collect teacher data (PASS2, instructional surveys, interviews). • Collect student data (YouthTruth survey, NWEA MAP test in sci, reading, language) - exploratory. • Collect/analyze implement data. • Interview hubs and share feedback. • Provide implementation feedback to MSSL program staff.	• Eval meet with hubs to iteratively refine model. • Hub leaders meet quarterly to adjust plans, share strategies, and build network.

	Teaching & Learning	Research & Evaluation	Capacity Building
Y4	• TLs/content partners lead MSSL for 300 Ctrl Ts (phy sci), plus PLC (online). • Lead MSSL for 200 rural Ts (phy sci), plus PLC (online). • Collect digital library resources.	• Collect TL data (survey, interv). • Collect teacher data (PASS3, instructional surveys, interviews). • Collect student data (YouthTruth, NWEA MAP) - confirmatory. • Collect/analyze implement data. • Interview hubs and share feedback. • Provide implementation feedback.	• Hubs are self-sustaining with leadership training and sufficient resources to support ongoing PD. • Develop toolkit to support new hub leaders. • Establish 2 new hubs to support rural regions.
Y5	• TLs/content partners lead MSSL for 300 Ctrl Ts (Earth), plus PLC (online). • Lead MSSL for 200 rural Ts (Earth sci or hub choice K-8 Ts) • Expand online PLC availability to new hubs. • Expand use of digital library to support NGSS CCSS nationally.	• Interview new/old hub members. • Provide implementation feedback. • Analyze feasibility and fidelity of implementation data. • Write comprehensive evaluation reports summarizing results. • Present findings on webinars and at professional meetings. • Publish journal articles.	• Hubs are self-sustaining with leadership training and sufficient resources to support ongoing PD. • Develop action plan for national scale up based on lessons learned. • Identify regions for expansion of MSSL hubs.

Qualifications of Key Personnel. This i3 project team is comprised of experts dedicated to improving science and literacy teaching and learning. Each offers unique expertise in professional development, technology, leadership, scaling, and/or research. (See Appendix F.)

Dr. Steve Schneider, STEM Program Director at WestEd and former PI of the WWC, will serve as PI, contributing expertise and guidance to research, management, fiscal planning, dissemination, and scaling. He will serve as liaison between the research team and program staff, act as point of contact to U.S. DOE, and lead communications with state/national organizations, drawing on established partnerships with influential foundations and corporations. **Ms. Kirsten Daehler**, Senior Project Director at WestEd, will serve as co-PI and Director of Programs. In this role she will oversee the day-to-day work, ensure tasks are completed on time and within budget, communicate with districts and site coordinators, and provide guidance to hubs. She will chair the Core Leadership Team comprised of representatives from all partners, who via monthly conference calls monitor project efforts. **Dr. Denis Newman**, Chairman and CEO at EEI, and **Dr. Joan I. Heller**, Director of HRA, will

also serve as research co-PIs. Together they and their staff, including **Dr. Andrew Jaciw (**Senior Scientist at EEI), will carry out all aspects of the i3 evaluation efforts, including quarterly reports and accountability information. These researchers have extensive experience conducting formative, process, and large-scale, experimental evaluations. Currently, EEI is conducting several evaluations for i3 winners.

A distinguished group of **Program Advisors** and a **Technical Working Group (**TWG), will attend annual meetings and provide ongoing phone and email consultation. Program advisors will help refine the implementation model, help plan for identifying and training teacher leaders), and review prototypes of the online portal. The TWG will help to refine the research design, contribute to instrument development, and make recommendations to guide data collection and methods of analysis. (See Table 3.)

Table 3. *Program Advisors and Research Advisors.*

Program Advisors	Research Advisors
• *Dr. Christopher Dede,* Wirth Professor in Learning Technologies, **Harvard University** • *Kathy Dunne,* PD, **WestEd** • *Dr. Elyse Eidman-Aadahl,* Co-Director, **National Writing Project** • *Dr. Cynthia Greenleaf,* Co-Director, **Strategic Literacy Initiative** • *Ms. Deborah Levitzky,* Founding Director, **National Academy of Advanced Teacher Ed**	• *Dr. Hilda Borko,* Professor of Ed, **Stanford University** • *Dr. Anne Chamberlain,* Senior Research Associate, **IMPAQ International** • *Dr. Kathy Comfort,* Director, **Partnership for Assessment of Standards-based Science** • *Dr. Heather Hill,* Professor of Education, **Harvard Graduate School of Education** • *Dr. Ellen Kisker,* Senior Research Scientist and President, **Twin Peaks Partners, LLC**

Program Leads, Ms. Jennifer Folsom and Ms. Jo Topps from WestEd, will train MSSL facilitators and support the cadre of teacher leaders. **Technology Leads,** Ms. Patricia Wasley, CEO at Teaching Channel, and Mr. Robert Montgomery, will oversee development, testing, and refinement of the Tch's portal, digital library, and online PD tools.

Multi-year Operating and Financial Model. Major programmatic expenses for this ambitious five-year project have been carefully planned and budgeted to ensure that resources are adequate

to meet project objectives. For example, our financial model includes incentives for research activities, stipends for participation, staff time, and compensation for TLs and site coordinators, as well as travel associated with research and dissemination activities to increase participation and retention in the research. The streamlined operating budget reflects WestEd's experience in adequately allocating resources for large-scale, multi-year efforts, and builds on the PI's work as director of the IES National Center for Cognition and Mathematics Instruction ($10M) and the NSF Center for Assessment and Evaluation of Student Learning ($12.2M).

As a large nonprofit organization, WestEd's operating model designates directors from its contracts and finance departments to work closely with i3 project directors to systematically monitor expenses and provide formative data to assist in budgeting for and managing the work. Monthly financial reports compare projected costs with actual expenditures, and include multiple reports on staff allocations, as well as costs over time and by category, which provides the data for project directors to adjust plans, as needed, to meet project goals on time and within budget.

The operating and financial model for MSSL capitalizes on a regional hub affiliate model that provides direct and close-to-the-customer services while being supported by WestEd's national (centralized) infrastructure. For example, existing hubs such as the Discovery Place Education Studios (NC) and Los Alamos National Labs (NM) meet the specific needs of their local teachers, while implementing high-quality MSSL courses. To build regional hub capacities, the plan will support training and certification of providers, along with structures for information-sharing across hubs while feeding data back from WestEd to support iterative refinement of regional hubs. Prior work has provided useful budget information, such that start-up costs are known and a proof of concept has established a viable cost-recovery model of providing MSSL to districts. We have confidence that the budget is allocated to support the work because it is based directly on historical budget data from prior work.

QUALITY OF THE PROJECT EVALUATION

Evaluating Goal 1: Transforming Teaching and Learning.

Successful PD scale-up requires impact, depth, sustainability, and spread (Coburn, 2003). The evaluation will investigate these dimensions through: (a) a *student and teacher impact study* to validate effects of the scaled-up MSSL program on teacher outcomes and student science achievement, and (b) a *regional capacity-building study* that will evaluate the scale-up model within the collaborative hubs in CA and WI. For this evaluation, EEI will manage and conduct a randomized controlled trial to assess the impact of the project on teaching and learning (Goal 1), and will also assess the project's cost effectiveness (Goal 3). HRA will evaluate the project's strategies for scaling up MSSL (Goal 2). Evaluators will also provide quarterly formative feedback to project staff through documented, in-person briefings and provide hub leaders with progress reports about strengths and needs of implementation efforts.

Study questions and methods. The impact study to evaluate the MSSL intervention will apply mixed methods to assess the key components of the logic model, including presence of inputs (e.g., teacher leadership development); impacts on proximal and intermediate outcomes (e.g., improved content knowledge of teachers); impacts on distal outcomes (e.g., improved student achievement); and mediating effects on the distal outcomes.

Table 4: *Overview of impact study schedule.*

Phase	School Year		Grade Level			Research Activities
		2015	3A			Planning; recruiting and randomization
Capacity-building	2015	2016	3B	4A	5	Formative evaluation; pilot of class observation protocol and student survey
Full implementation of scale-up and impact study	2016	2017	3C	4B	5A	Evaluation of exploratory impact; field test class observation protocol and survey
	2017	2018	3	4C	5B	Evaluation of confirmatory impact; cost benefit and start of rural scale-up study
Expansion to rural schools	2018	2019	3	4	5C	Analysis of confirmatory impact ; Evaluation of rural scale-up
	2019				5	Analysis and reporting

A	Cohort A	**B**	Cohort B	**C**	Cohort C		Assessment of confirmatory outcomes

Preliminary planning, participant recruiting and randomization happen in Spring 2015, while leadership development of a cadre of teachers in treatment schools, and instrument refinement, run through 2015/16. During "full implementation" (2016/17 and 2017/18), teachers in grades 3–5 in treatment schools will participate in MSSL activities. Achievement outcomes for the primary impact analyses will be collected in Spring 2018 in 4th and 5th grade. (See Table 4.)

Goal 1 Research Questions. Generated from our theory of change, the primary research question is: (a) What is the impact of broad adoption of MSSL, following two years of full implementation on 4th and 5th grade students' science achievement and literacy, as measured using NWEA MAP tests of science content knowledge, reading, and language usage? Secondary research questions are: (b) What is the impact of broad adoption of MSSL on: (b1) Teachers' science content knowledge as measured through a PASS assessment? (b2) Teachers' teaching practices and students' discourse patterns, as measured through the Science Quality of Instruction (SQI) classroom observation instrument, an adaptation of the Mathematical Quality of Instruction (MQI)? and (b3) Students' dispositions and attitudes towards science and future science learning as measured through the YouthTruth Elementary School Survey?

In addition, researchers will conduct a series of exploratory analyses to more fully understand the variation in impact and to assess mediating mechanisms. Exploratory research questions are: (c) What are student, teacher, and school-level attributes that *moderate* impacts of MSSL on student achievement? For example, researchers will explore the hypothesis that impact of MSSL on science achievement is greater for English language learners and minorities, thereby reducing the achievement gap, a result observed in prior impact evaluations of MSSL. Other potential moderators include school factors such as socioeconomic status, and proportion of Limited English Proficient students. (d) Do impacts of MSSL on teacher effectiveness in science instruction or content knowledge *mediate* impacts on student achievement in science?

As part of exploratory analyses, researchers will also assess impacts after one year of full implementation, whether impacts are correlated with levels of implementation, and whether impacts vary across randomized blocks, and if so, which moderators account for this variation.

Design, sampling, power analysis, and random assignment.

Design. This three-year cluster randomized trial will be designed to meet high levels of internal validity (*WWC standards without reservation*) and statistical conclusions validity. The design attempts to limit potential compromise of validity through contamination (randomization is of schools instead of within schools) and attrition (obtaining teacher and district commitment, offering incentives). Further, measures will be utilized that have strong reliability and validity and that target the critical constructs without being over aligned with the intervention.

Sampling. Sixty schools will participate in the experiment, with 30 schools drawn from each of WI (Milwaukee) and CA (San Joaquin). This diverse sample will ensure efficacy is evaluated across a variety of contexts—ranging from urban to rural—and for a variety of students. Confirmatory impacts will be assessed in spring 2018 in grades 4 (in WI) and 5 (in CA). We expect 180 teachers (60 schools, with 3 teachers per school) and 4,500 students (25 per teacher).

Power analysis. The school sample size is determined by the primary research question focusing on impacts on students' science and literacy achievement. We rely on empirically-based estimates of parameters from studies with school randomization and science achievement outcomes (Newman et al., 2012; Westine, Spybrook & Taylor, 2014). The intraclass correlation coefficient (ICC) from both studies was .19. We set the school level R-squared to 0.70 (which accounts for effects of both blocking and modeling covariates including pretest). Confirmatory impacts on students will be assessed in 4th grade in WI and 5th grade in CA, for an average of three teachers (and 3 x 25 = 75 students) per school, and attrition rates of 10% and 20% at the school and student levels respectively. With power of 80% and assuming Type-1 error of 0.05

the Minimum Detectable Effect Size is 0.18, which is within the range observed in previous impact evaluations of MSSL and is smaller than the magnitude of impact previously observed at the elementary level. This result was obtained using the formula for MDES from Bloom, Richburg-Hayes, and Black (2007). Sixty schools should also provide sufficient power to assess impacts on student achievement in reading, for which unconditional ICCs have been consistently lower than for science (Westine, Spybrook, & Taylor, 2014). While impacts of inquiry-based science interventions on reading have been of similar magnitude as impact on science (Newman et al., 2012); reducing the unconditional ICC to .16 (the upper range for reading in Westine et al., 2014) lowers MDES to .17, with the rest of the parameters the same as above.

For impacts on teachers we focus on classroom observation outcomes. We select an IC of .21 which is the median value from studies of impacts on classroom observations (James-Burdumy, et al., 2009; Glazerman, et al., 2008; Schochet, 2009; Hough, 2010). Assuming R-squared values of .40 at the school and teacher levels, 3 teachers per school, attrition of 10% of both schools and teachers, power 80%, and Type-1 error of 5%, the MDES is .37, which is within the observed range for the studies noted above. This is lower than the range of impacts on teacher content knowledge observed in prior studies of MSSL, which spanned 0.4 and 2.0 standard deviations. (More details of the power calculations are provided in Appendix J.)

Random Assignment. Schools will be randomly assigned within randomized blocks to MSSL or control in Spring 2015 (30 in WI and 30 in CA). Schools assigned to the control condition will be wait-listed to receive the program in 2018/19. Impacts will be assessed in 4th grade in WI and 5th grade in CA (the grades at which science has traditional been evaluated) and analyzed together in Spring 2018 (i.e., see Cohorts B and C in Figure 6). Randomized blocks will be formed within each of the districts. Principal surveys, school-level demographic data from NCES and administrative records will be used to identify the blocks. Researchers will consider

geographic and demographic data (e.g., urbanicity, SES), and principal responses to baseline surveys about conditions that may be related to success in implementing MSSL (e.g., competing initiatives, levels of collaboration among teachers). Blocks of 3 or 4 schools each will allow modeling variation in impact across blocks, and moderators of the variation.

Measures and Data Collection.

Background. *Principal surveys* will be administered once prior to randomization concerning factors potentially predictive of achievement outcomes and uptake of MSSL, such as levels of commitment to science and literacy instruction and levels of collaboration. *Teacher surveys* will be completed at baseline and then once every year over the course of the trial. Questions will address attitude/confidence towards teaching, and opportunities they provide students to learn science (hours per day/week) and to read, write, and talk about science. Teachers will also be asked about their background (e.g., total years teaching, degrees, credentials, coursework).

Impacts on students. *Northwest Evaluation Association (NWEA) Measures of Academic Progress* (MAP) tests of Reading, Language Usage, and Science will be used (for the CA and WI content aligned tests, test-retest correlations for reading, language, and science range between .67 and .81; validity of concurrent performance on state tests ranges between .77 and .82). MAP for Science covers specific concepts within three major domains of science: life, earth and space, and physical science, and thus is aligned with the MSSL domains in grades 4 and 5. NWEA science and reading tests will be used to establish individual baseline scores.

The *YouthTruth student survey* will be used to assess engagement in science and disposition and attitudes towards science and future science learning. The survey includes subscales addressing Student Engagement, Academic Rigor and Expectations, and Instructional Methods (coefficient alphas range 0.82-0.93), as well as Student Attitudes toward STEM Learning. (See Appendix J.) Researchers will partner with YouthTruth to pilot, field test, and fully implement

additional subscales to capture constructs pertaining to experiences of effective science instruction and attitudes towards future science learning. The survey will be administered on a limited scale during the pilot phase (2015/16, 2016/17) and with the full sample in 2017/18.

The measures described are strong options. However, because science assessment is in a state of great change we will continue to investigate alternatives. For example, we are in conversation with the Educational Testing Service (ETS) concerning their CBAL (Cognitively Based Assessment *of*, *for*, and *as* Learning) initiative, which will include a science assessment.

Impacts on Teachers. The *PASS (Partnership for the Assessment of Standards-based Science)* test will be used to assess teacher content knowledge. PASS, which was developed with support from the NSF, consists of a suite of standards-based selected response items, constructed response items, and hands-on performance tasks (see Appendix J). The reported score reliability is 0.87, and inter-rater reliability between 0.84 and 0.87. PASS can be customized to focus on individual topics, making it a strong option for measuring impact. The assessment will be administered in teacher meetings each spring preceding the summer PD institutes.

Science Quality of Instruction (SQI) is an observation protocol that will be adapted from the Mathematics Quality of Instruction (MQI), with permission from and in consultation with advisory board member and MQI developer, Dr. Heather Hill (personal communication, June 2, 2014), These dimensions include the richness of the content, student reasoning and meaning-making, and the clarity and correctness of the content. These dimensions overlap substantially with the MSSL model, and no instrument exists for assessing science instruction that approaches the extensive research base on validity and reliability of the MQI. Studies suggest reliable teacher scores, G-study reliability of .77 (Learning Mathematics for Teaching, 2011), and those scores correlate with student outcomes (Hill et al., 2008; Hill, Charalambous, & Kraft, 2012). As needed, modifications will be made based on existing literature and validated through analysis of

a sample of 40 videotapes. EEI has collected classroom video during a previous i3 Validation grant and has the camera/microphone equipment, video-streaming software and coding tools, and hosting service. Building on this experience, two video rigs will be placed in the classroom, one in the middle and one on the front board. We plan to conduct observations with all 180 teachers in MSSL and control schools during 2017/18 and observe four consecutive lessons per teacher. We will also pilot the process in 15 schools (45 teachers) in spring 2016 and 2017.

Analysis Plan

Student outcomes. Analyses address the primary research question concerning impacts of MSSL on science, reading, and language achievement after two years of full implementation, as well as secondary research questions concerning impacts on these outcomes after one year of full implementation, and on student attitudes and dispositions towards science. Intent-to-treat (ITT) estimates of impacts on student achievement outcomes will be obtained using hierarchical linear models (HLM) (Raudenbush & Bryk, 2002; Singer, 1998) as applied to cluster randomized trials (Bloom, Bos, & Lee, 1999, 2005). The distribution of each scale (normal, censored, count, or binomial) will be determined to select the most appropriate model. The benchmark impact model will include dummy variables to reflect randomized blocking, a dummy variable to indicate treatment status, a school-level random effect to account for clustering of students in schools, and student- and school-level covariates. Covariates will include the school-level science pretest for 4th and 5th grade cohorts collected in Spring 2015 (before randomization) and individual NWEA pretests. Other covariates will likely include LEP status, socioeconomic status, and ethnicity. The dummy variable method will be used to address missing values for covariates (Puma et al., 2009), and cases with missing outcomes will be listwise deleted. Teacher-level random effects will not be included since doing so with school-level randomization has limited influence on the impact estimate or its standard error (Zhu et al., 2012).

Teacher outcomes. Analyses address the secondary research question concerning impacts of MSSL on teachers' content knowledge and teaching after two years of implementation. The impact model will be similar in form to the one used for student outcomes—a two-level model—with school- and teacher-level covariates, dummy variables for randomized blocks, and a school-level random effect. (The HLMs for assessing impacts on students and teachers are provided in Appendix J.) For teacher knowledge and classroom observation outcomes, in addition to the baseline teacher knowledge total scores, the model will include total teaching experience, teacher's education level (master's degree or not), science major or not, and number of postsecondary science courses taken.

Classroom teaching. Videos of teaching will be coded by a team of EEI employees trained to use the SQI. As video observations are collected, the coding team will pilot the protocol with a subset of the collected video. Raters will be trained on the protocol and a reliability test will be conducted before general coding begins. Raters will be asked to code a specific number of classroom segments. Their codes will be compared with those provided by expert raters from the MQI team. The target reliability is 80% exact agreement across the classroom segments. Videos will be coded by raters using a computer interface designed by EEI engineers. This interface will provide raters with a queue of videos generated by algorithm. The coding team will be unaware of treatment condition and other possibly biasing information about the teacher and school. Most observations will be coded by one person, but 10% of videos will be coded by multiple raters for use in "calibration conversations" and reliability checks.

Exploratory analyses. Differential impacts will be assessed by adding a term for the interaction between the indicator of random assignment status and the hypothesized moderator. The interactions will be assessed individually, not simultaneously in one model. Researchers will examine the moderating effect of student population variables associated with differential

academic achievement, individual-level pretest, socioeconomic level of the school population (percentage of students eligible for free or reduced-price meals), and percentage of students who are Limited English Proficient, as well as science-related resources at the school, and most importantly, presence of science kits or other materials needed for science investigations. Mediator analyses will be conducted within a multilevel framework (Krull & MacKinnon, 2001). Because with the regression-based approach to mediation, the relation between the outcome variable and the mediator may be confounded by other relations, we will also use a principal stratification approach (Frangakis & Rubin, 2002; Jo, Stuart, MacKinnon & Vinokur, 2011; Page, 2012). The mediation analyses will help to determine the active paths in the logic model, thereby indicating where to invest in future iterations of the intervention, potentially resulting in more benefit for less cost. Levels of fidelity will be assessed through a numeric index defined through careful consultation with the program developers and based on the qualitative work.

Researchers will use approaches for estimating impact under conditions where the program is adequately implemented (Unlu, Bozzi et al., 2010) that build on the literature on principal stratification. They will explore Complier Average Causal Effect (CACE) analyses (Angrist, Imbens et al., 1996; Frangakis, Rubin et al., 2002) as an alternative to ITT, and Local Average Treatment Effect (LATE) analyses to address the problem of crossovers should it arise (Gennetian, Morris, Bos, & Bloom, 2005). Attrition and differential attrition will be monitored at each level of the design and the potential for bias assessed using *What Works Clearinghouse* standards. Researchers will also conduct a series of sensitivity analyses to test robustness of benchmark impact estimates for the primary research question; including analyzing impacts at the cluster level, by modeling impacts as randomly varying across randomized blocks, and using different approaches to addressing missing values, such a multiple imputation, and simple listwise deletion. Researchers will also produce an estimate where impacts from the two states

are explicitly given equal weight. Analyses will be conducted using PROC MIXED and GLIMMIX in SAS as well as specialized programs such as *Rmediation* (Tofighi & MacKinnon, 2011) and *mediation* in R (Imai, Kosuke et al., 2010).

Evaluating Goal 2: Enable Broad Adoption of Effective Practices.

The long-term goal of this project is to create organic, self-sustaining regional hubs and a system of teacher leadership development that support broad adoption and ongoing delivery of teachers. The evaluation will consist of a fidelity-of-implementation study to determine whether the program delivers critical components (inputs and activities in the logic model in Figure 4, p. 10) consistent with the essential practices of the PD intervention and scale-up model (Figure 5, p. 14). The study will also investigate, in the context of all hubs established in Year 1, whether the scaled-up implementation in Years 2–4 shows promise for being sustained after the project ends.

Goal 2 Research Questions. Questions for this goal are: (a) To what extent and in what ways do the regional hubs result in collaboration among stakeholders to provide resources and services to schools and districts? (b) To what extent and in what ways do the regional hubs increase teacher leaders' capacity to deliver science professional development? and (c) What promise do the regional hubs show for continuing to increase leadership capacity and support delivery of science professional development going forward? (d) What important barriers and supports for building regional capacity were encountered over the course of the project?

These questions will be addressed by monitoring key indicators throughout the project, identifying thresholds for adequate implementation, and collecting the same data each year to track changes over time. (See Table 5.) We will collect data from: hub coordinators, leaders of participating organizations, PD facilitators and teacher leaders, and teachers in all grades included in hub activities. Both objective and self-report data will be collected to describe and document events, resources and services delivered, and participation in project activities.

Table 5. *Key Variables, Data Collection Methods, and Metrics in Goal 2 Scale-Up Study.*

Variables	Data collection method	Metric
Broad adoption of high-quality with fidelity to model.	Facilitator academy & PD/PLC records of attendance in Y2–5. Post-PD written surveys of all teachers in Y2–4. Online facilitator logs, observations of PD.	Number of MSSL facilitation academies. Number of MSSL teacher leaders trained. Number of MSSL courses offered. Numbers of teachers & students served. Completion of PD session components. Teacher reports of PD features.
Use of online tools and national library of digital resources.	Annual usage summaries. Online surveys of teacher participants in winter and spring in Y2–4.	Teaching Channel usage data, number of artifacts in digital library, discussion forums. Self-reports of use and value of available technologies.
Collaboration of stakeholders that increases teaching resources.	Brief bimonthly online surveys of hub participants in Y2–5. Interviews with hub coords winter & spring Y2–5.	Number of meetings and conference calls among hub participants. Reports and verification of services and resources contributed and received.
Teacher leaders' reach extended to rural and other isolated teachers.	Focus group phone interviews with participating teachers. Observation of online PD sessions.	Numbers of teachers & students served in rural and outlying areas. Differentials in services provided, teacher retention, teacher reports of PD process/value.

As part of the implementation study researchers also will assess the achieved relative strength of the intervention-control contrast (Cordray & Pion, 1993; Cordray & Hulleman, 2009) by measuring, through surveys, the degree to which the intervention model, in practice, differs from the pedagogical model(s) underlying the business-as-usual comparison condition. For example, teachers in both conditions will be surveyed concerning the amount and characteristics of the science and literacy PD they receive. This will allow interpretation of impacts in the context of achieved PD implementation in both conditions. In addition to investigating the presence of specific inputs and activities in the implementation study for accountability purposes, we will investigate potential impediments to and mediation of impact.

This i3 validation is well situated to make important contributions to improving student achievement, impacting high-needs urban and rural communities, furthering research, and building regional hubs with the expertise and materials needed for sustainability.