_L_iteracy-_I_nfused _S_cience Using _T_echnology Innovation _O_pportunity (LISTO)

Table of Contents

Section **Page**

***L**iteracy-**I**nfused **S**cience Using **T**echnology Innovation **O**pportunity (LISTO)*

Texas A&M University's (TAMU) Center for Research and Development in Dual Language and Literacy Acquisition (CRDLLA), Education Leadership Research Center (ELRC), and Aggie STEM, in collaboration with 70 Texas schools and Johns Hopkins University (JHU; external evaluators), proposes *L*iteracy-*I*nfused *S*cience Using *T*echnology Innovation *O*pportunity (LISTO), a 5-Year Longitudinal Validation Project under *Absolute Priorities* 2 (Promoting STEM Education) and 4 (Serving Rural Communities). The term, "LISTO," in Spanish means *Ready*, and we are ready to assist teachers and principals in building instructional capacity and improving students' science and reading/writing literacy achievement. Appendix G includes 37 LEAs' letters of support and commitment to participate under Project LISTO. The *overarching goal* of LISTO is to validate, via a longitudinal randomized controlled trial (RCT) study, literacy-infused science (LIS) instructional and curricular innovations from a prior research grant in order to increase instructional capacity of teachers and to improve students' science and reading/writing literacy achievement in rural/non-rural schools for economically challenged (EC), inclusive of English language learners (ELL) students [mainstream EC students, EC ELL, and EC or ELL students with disabilities who are inclusion students]. The prior grant, Project Middle School Science (MSSELL, DRL-0822153), was funded by the National Science Foundation; Tong, Irby, Lara-Alecio, Guerrero, Fan, and Huerta (2014) and Lara-Alecio, Tong, Irby, Guerrero, Huerta, and Fan (2012) (Appendix D) defined literacy-infused (literacy-integrated) science as reading to learn in science with specific reading/writing skills embedded in instruction and curriculum. LISTO is linked directly to the Next Generation Science Standards (2012) which state, "students must read, write, view, and visually represent as they develop their models and explanations. They speak and listen as they present their ideas or engage in reasoned argumentation with others to refine their ideas and reach shared conclusions" (p. 3).

LISTO will be implemented under new conditions with promising strategies and proven exceptional approaches, specifically, (a) across rural (46 or 66%) and non-rural schools (24 or 34%) with 560 teachers in Texas who are representative of those serving large numbers of needs middle school (11,200) students, including EC students and ELL student populations; (b) with treatment science classes compared to control/comparison science classes; (c) to determine the degree of impact of innovations on EC and ELL students' science achievement and reading/writing literacy skills, and (d) to facilitate scalability across a broad geographic region by using technology to bring the innovations to a variety of settings and populations. LISTO is based on *moderate evidence of effectiveness* from Project MSSELL (Appendix D; Tong et al. [2014] vetted by *WWC as without reservations*).

LISTO has a two-tiered sustainability validation model. ***Tier 1*** is a longitudinal design following the same 5th-grade (GR.) teachers – 140 teachers (2 teachers per each of 70 schools) will be included with four GR. 5 cohorts over 4 academic years with a total of 11,200 students in GR. 5 (140 teachers x 20 students per class = 2,800 students x 4 years). ***Tier 2*** is a longitudinal design following the same GR. 5 students over 4 years. Cohort 1 of 2,800 GR. 5 students who participated in the first year of implementation will be followed as they matriculate through GRs. 6, 7, and 8. Across those 3 years for Cohort 1 participants, their treatment (T) and control (C) teachers (anticipated 140 additional teachers per year) will receive virtual professional development (VPD) and virtual mentoring and coaching (VMC) innovation on how to infuse literacy into science, along with science content. In this way, teacher effect in GRs. 6-8 between T and C can be accounted for, and the only difference, if any, at student level by the end of GR.

8 would be due to the sustained impact of the full intervention that occurred in GR. 5 in the T schools. This two-tiered model is depicted in Figure 1.

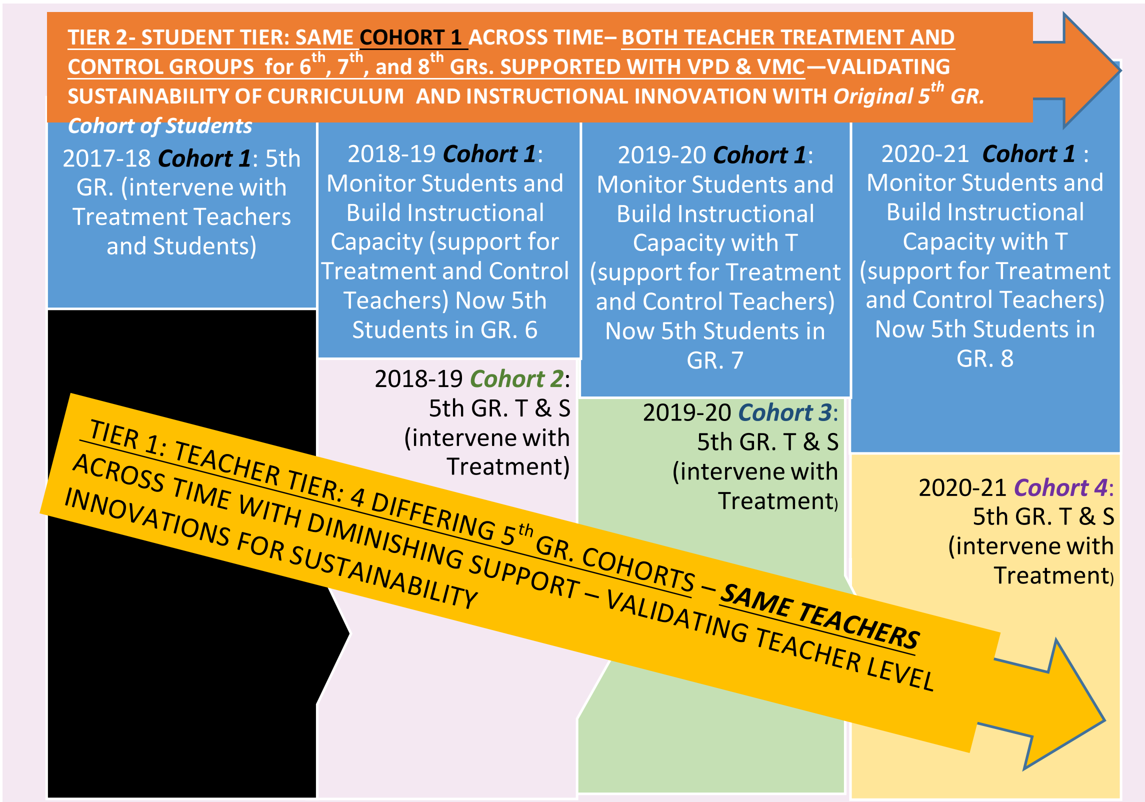

Figure 1. LISTO two-tiered sustainability validation model.

Absolute Priority 2 — Promoting STEM Education. LISTO will validate STEM education — science and technology innovations as indicated. **Absolute Priority 4 — Serving Rural Communities (from the RLIS program).** Two-thirds of LISTO's student participants will be located in 46 Texas rural schools across the state (RLIS-eligible districts). These rural districts serve large numbers of EC students (50.1% to 92.9%) and are inclusive of ELLs (3.2% to 68%) on the campuses. An overview of rural districts (Appendix J.1.) characteristics is listed in Appendix J.2.

A. Significance: A.1. Magnitude or Severity of the Problem to be Addressed

The magnitude of severity of the problem to be addressed in LISTO rests in five major issues with imbedded or related implications: (a) disparities between rural and non-rural schools, (b) major increases in the numbers of EC and ELL students and their academic challenges in all school types, (c) the need to build instructional capacity for in-service teachers in science and technology while addressing the learning needs of EC and ELL students, (d) the necessity of cultivating student interest early in STEM, particularly in science and in rural schools, and (e) lack of classroom observation studies for EC/ELLs.

Disparities Between Rural and Non-rural Schools — Almost half (49.9%) of the nation's school districts are small rural districts, and 32.9% of public school campuses are rural (Johnson, Showalter, Klein, & Lester, 2014). Four states with the largest rural enrollment, in order — *Texas*, North Carolina, Georgia, and Ohio — serve more than 25% of all rural students in the nation (Johnson et al., 2014). The research *on teacher quality and the achievement gap in rural schools has not been widely studied as compared to teachers in urban and suburban districts* (Cowen, Barrett, Toma, & Troske, 2015). **Teacher quality.** Due to difficulties with the recruitment and retention of highly qualified teachers (Webb, 2006-07), particularly in the *STEM* subjects (Ossola, 2014; Monk, 2007) in rural schools, students have limited access to teachers with content-area expertise (Beesley, 2011; Friedrichsen, Chval, & Tuescher, 2007; Monk, 2007). Further, rural schools face obstacles delivering effective professional development (PD) to teachers due to distance, and lack of PD resources and personnel to support PD (Glover et al., 2016). **Targeted research on rural schools and science.** Eight published papers appeared in a major search of TAMU library databases on *rural schools and middle school science and STEM*. The research included the following topics: (a) two on science and writing, (b) three on gender and attitude toward and/or achievement in science/technology, (c) one on fast-food consumption, (d) one study in Turkey, and (e) one on relationship between student and teacher.

Increases in and Challenges for EC and ELL Students — Students who are eligible for free or reduced-price meals under the National School Lunch and Child Nutrition Program are identified as economically disadvantaged, or as we use the term, *economically challenged.* Approximately 20% of U.S. school-age children live in poverty, and 25% of public schools are considered high-poverty schools (Kena et al., 2016). Texas has a higher rate of childhood poverty than the national average (Kena et al., 2016), with 14% of all its districts (164) falling under the Rural and Low-Income School (RLIS) Program (U.S. Department of Education, 2015). The Texas Education Agency (TEA, 2016) reported that the rate of enrollment of EC students grew from 54.5% in 2004-05 to 58.7% in 2014-15, and the increase in EC students has outstripped the overall growth in student enrollment. On the other hand, ELLs are those students whose first language is not English and whose limited English language skills make it difficult for them to complete classwork in English (TEA, 2015). The percentage of students identified as ELLs increased from 14.4% in 2004-2005 to 18.1% in 2014-15 (TEA, 2015). ELLs represent 3.1% of rural students (Johnson et al., 2014). According to TEA (2015), for 2014, EC students made up 60% of students who completed State of Texas Assessments of Academic Readiness (STAAR) assessments in GRs. 3-8. The average passing rate for EC students on all tests in GRs. 3-8 was 65%, which was lower than the state average for all students (74%) (TEA, 2015). For those ELLs found to be at risk, their passing rates fell from the elementary to secondary levels, with a high of 66% in GR. 3 to a low of 35% in GR. 7; similar patterns were also observed for monitored or former ELLs at risk (TEA, 2015). Furthermore, although acquiring scientific literacy presents a challenge to all students (Gee, 2005; Ryoo, 2010), those who are of color, who come from EC backgrounds, and/or those who are ELLs are presented with the greatest challenge. For example, according to the latest data from the 2011 National Assessment of

Educational Progress (NAEP), the percentage of ELLs at or above proficient at GR. 8 was 2% in science, as compared to 32% in science among English speakers; for EC students, 44% were below basic in science, while 73% of ELLs were below basic (National Center for Education Statistics [NCES], 2012a). Such is also reflected STAAR in which GR. 5 and 8 (GR. levels assessed) ELL and EC students had lower passing rates (ELLs: 52%, 38%; EC students: 63%, 62%) compared to overall state passing scores (72%; 71%) (TEA, 2016). **Literacy and science.** We found few studies addressing literacy and science instructional integration at the middle school level for ethnically diverse, low-SES students (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009; Lara-Alecio et al.,, 2012; Palumbo & Sanacore, 2009). Furthermore, from a methodological perspective, experimental and quasi-experimental designs with such students are rare; few studies are longitudinal that result in positive intervention effect on student achievement (August et al., 2009; Lee & Luykx, 2006). Finally, Stoddart, Pinal, Latzke, and Canaday (2002) indicated there was a limited availability of published research on any type of integration of science, particularly inquiry-based science curriculum, with reading for EC students and ELLs; we determined this is still the case in 2016. We found no studies on EC/ELL students in rural schools pertaining to LIS curriculum and instruction.

Building Teacher Instructional Capacity — According to Byrnes, Kiger, and Manning (1997) and Lee (2005), most classroom teachers have had minimal, if any, training in meeting the academic or linguistic needs of their ELLs, and, in fact, McCloskey (2002) reported that only 12% of teachers nationwide had any training on how to teach with ELLs, much less when the content area of science is added. Even with some PD on strategies to make content comprehensible for EC students and ELLs, mainstream teachers did not accommodate the students' learning needs as they should (Brown & Bentley, 2004; Lara-Alecio, Tong, Irby, &

Guerrero, 2013). Although Thadani, Cook, Griffis, Wise, and Blakey (2010) and Amaral et al. (2002) addressed the equity issues among EC students and ELLs through curriculum-based interventions in science education, there were no large scale RCT studies, and none in rural schools. Literacy is another crucial area for teacher PD, but few middle school teachers receive PD on integrating literacy instruction into content (Heller & Greenleaf, 2007). As a result, they are unfamiliar with research-based reading practices and are hesitant to or do not know how to implement such strategies (Kamil, 2003; Ness, 2007), and furthermore, only 10% of teachers give students an opportunity to experience science in an interactive way (Exploratorium, 2016).

Student Interest in Science — EC students are less likely to enter STEM fields than their peers from more advantaged family backgrounds (Chen & Weko, 2009). Potvin and Hasni's (2014) systematic literature review indicated that student interest in science declines over the course of K-12 education, especially at the point of the elementary/secondary transition. The authors noted that this drop is sharper in *rural and pedagogically traditional* classrooms. However, many school factors influence student interest in science, including enthusiastic, engaging teachers, use of hands-on, inquiry-based learning, and laboratory experiments (Potvin & Hasni). Teachers are key to nurturing student motivation and self-efficacy in science (Redding & Walberg, 2012).

Research on motivation. Researchers have indicated a relationship between student achievement and motivation (Ames, 1990; Liu, Horton, Olmanson, & Toprac, 2011); some scholars noted this relationship is stronger for marginalized students than other groups (Woolley, Strutchens, Gilbert, & Martin, 2010). **Developing a science identity.** By the age of 13, most students have made decisions about what they do not want to do in a future career, a decision that influences their career goals into high school and college (Extraordinary Women Engineers Project, 2007); this is why it is critical to work with students around GR. 5 — 11 years old.

Classroom Observation Studies with EC/ELLs — While extensive literature has been devoted to systematic observation in English-only classrooms (see Brophy & Good, 1974; Stallings, 1980; Waxman, Huang, Anderson, & Weinstein, 1997; Waxman, Rodriguez, Padrón, & Knight, 1988), only a few studies have been conducted exclusively with ELLs (e.g., Breunig, 1998; Brisk, 1991; Bruce et al., 1997; Escamilla, 1992; Irby, Tong, Lara-Alecio, Meyer, & Rodriguez, 2007; Padrón, 1994; Ramírez, Yuen, Ramey, & Pasta, 1991; Strong, 1986). As Lara-Alecio et al. (2009) pointed out, only a few studies rest on reading/language arts (e.g., Foorman & Schatschneider, 2003), content reading in science and social studies (e.g., Irby et al., 2007), or a mixture of reading/language arts (e.g., Foorman et al., 2006). Gersten and Baker (2000) claimed that little is known on how instructional intervention can shape teachers' pedagogical delivery. Garza (2015) indicated a lack of reliable and valid classroom observation instruments, and therefore, the corresponding empirical evidence masks the actual classroom practice with teachers who work with ECs and ELLs.

A.2. & A.3. Promising New Strategies and Exceptional Approach to Competition Priorities

LISTO is aligned to the Texas Equity Plan that sets forth high-impact practices that TEA will take to ensure that poor and minority children are not taught at higher rates than other children by inexperienced, unqualified, or out-of-field teachers. LISTO provides high-impact, high-feasibility innovations for (a) targeted PD for participating teachers, (b) mentoring and coaching for participating teachers, as well as compensation, (c) PD for participating campus leaders, and (d) rewards (Continuing Ed PD hours, at least 30/year.) and recognition in developed webinars.

As indicated in Figure 1, we shared two tiers of sustainability, and now we present the two *levels of innovation* in LISTO — instructional capacity for teachers and academic support for students — as indicated in Figure 2. The new conditions, promising new strategies, and exceptional approaches are: (a) new schools across Texas at 5th GR., (b) follow-up teacher PD and student monitoring through GR. 8, (c) student-engaged, standards-aligned curriculum that includes the development of academic language in science content (Tong et al., 2014); (d) a VMC (Irby, 2015) model (Appendix J.3.)with online delivery, in real time with no delay in feedback, (e) inexpensive technology for all teachers for VPD (Tong, Irby, & Lara-Alecio, 2015; Appendix J.4.; J.5.), (f) virtual observations (VOBS; Appendix J.6.) in the classroom with a platform for observing (Lara-Alecio, Irby, & Tong, 2015), (g) virtual science writing notebooks for Written and Academic oral language Vocabulary development in English in Science (WAVES; Huerta, Lara-Alecio, Tong, & Irby, 2014; Appendix J.7.), (h) Family Involvement in Science (FIS; Appendix J.8.; J.9) with virtual engagement and observation methods (Lara-Alecio et al., 2012), (i) Technology Infusion for English Literacy Advancement in Science

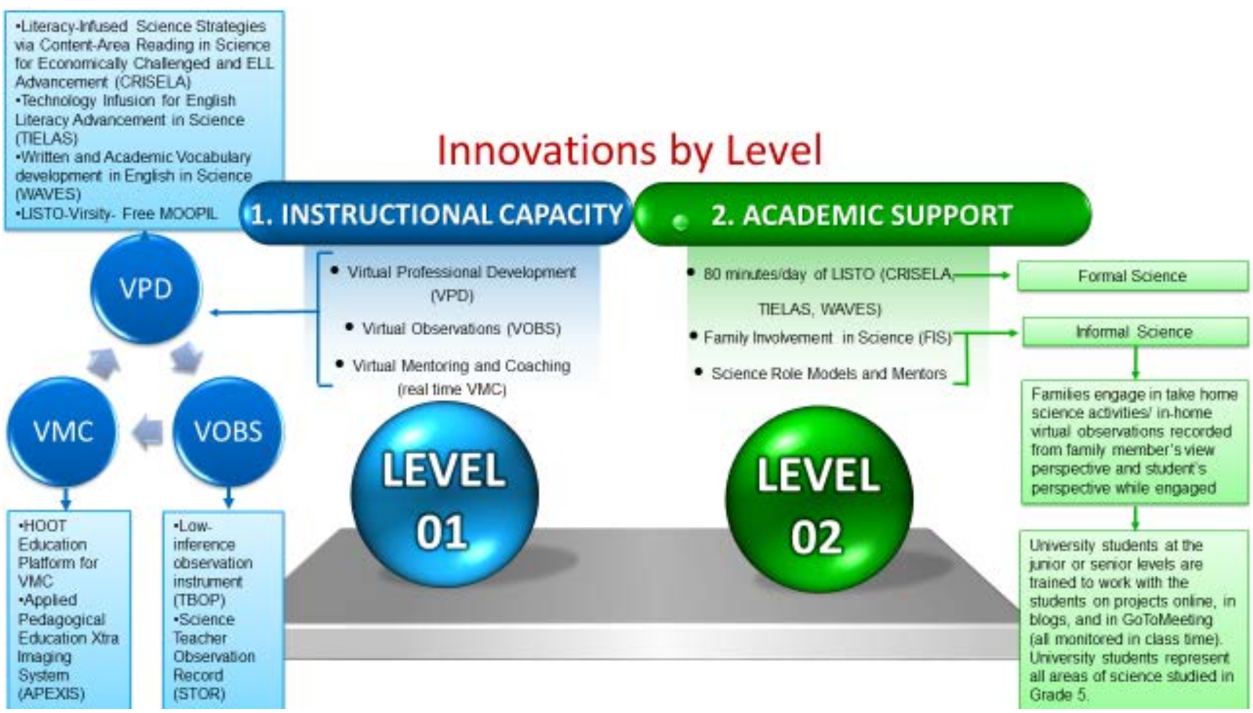

Figure 2. LISTO levels of innovation.

(TIELAS) with EduSmart Science with added higher order questions (Lara-Alecio et al., 2012; Appendix J.9.), and (j) Scientists as Role Models and Mentors (SRM[2], Lara-Alecio et al., 2012), which connects university science majors as mentors to grade-level students. The new conditions, promising new strategies, and exceptional approaches adapted from the original MSSELL and the successful virtual PD, mentoring and coaching, and virtual classroom observation from the current i3 validation grant (Project English Language and Literacy Acquisition—Validation (ELLA-V, U411B120047) are depicted in Appendix J.10.

LISTO will utilize a literacy-infused science (LIS) curriculum that supports reading skills using expository (informative) science text (Appendix J.11). Students will receive 80 minutes of daily science instruction for 26 weeks. The science curriculum, **CRISELA**, is strategically infused with instructional components to facilitate student reading, comprehension, and development of science concepts, following the 5E hands-on science model (Bybee, 1987). Teachers will provide direct instruction (pre-teach pronunciation of academic vocabulary, highlight tricky letter-sound combinations) and incorporate science academic vocabulary with student-friendly definitions and visuals, informative text features (e.g., headings, captions, text organization), strategic partner reading, and leveled comprehension questions. Students will participate in **WAVES,** using personal virtual notebooks (tablets/stylus) in which they process science content as they predict, record, organize, draw, question, and reflect. Treatment classrooms will benefit from **TIELAS** — equipping classrooms with instructional technology (e.g., teacher laptop, student tablets and as needed — projectors, document camera) and the integration of science educational software (EduSmart). Informal science learning is extended into the home with **FIS**, which will utilize point-of-view camera glasses to capture family/student interactions during researcher-developed home science activities. FIS activity packs, available in both English and Spanish when needed, outline science concepts students are learning, provide related vocabulary and definitions, link to related online games and/or science materials. Lastly, students will participate in **SRM^2**, a component designed to get students motivated about STEM. SRM^2 is a virtual mentoring program and blog involving university science majors.

Teachers will be provided with **VPD**. This will be delivered through high definition video conferencing (Citrix GoToTraining) with interaction via voice, chat logs, polls, and webcam. Sessions focus on student learning, instructional strategies, building capacity for science teaching, previewing upcoming lessons, viewing modeling videos, conducting inquiry activities, and reflection on student learning and teaching practices. Each session will be recorded and links will be sent out to participants so they can have access to go back and review. Teachers will also participate in **VMC** using our **APEXIS**) hardware platform and **Hoot Education (Hoot)**, online platform for teacher mentoring and coaching. A trained coach uses APEXIS to virtually observe teacher's instruction and offer support to treatment and immediate feedback. LISTO VMC offers support to teachers and an increase to fidelity of implementation. This approach allows teachers and coaches to have a voice to express their point of view. Teachers complete feedback based on reflection of their own instruction. Tiered coaching support is provided based on the level of fidelity from observations. The APEXIS hardware platform will serve a dual purpose so in addition to facilitating the communication for real-time VMC, it will be the platform utilized for conducting virtual classroom observations for the participating treatment and control teachers. **LISTO-Virsity** is defined as a Massive Open Online Professional Informal Individual Learning (MOOPIL; Irby, Sutton-Jones, Lara-Alecio, & Tong, 2015; Appendix J.12.) in which teachers will be able to access just-in-time PD online which ranges from 15 minutes one-time to PD up to four weeks. LISTO-Virsity PDs will be developed and supported by Aggie STEM, CRDLLA, and ELRC.

B. Strategy to Scale

Coburn (2003) suggested that scaling is a complex task of broad outreach "while simultaneously cultivating a depth of change" for support and sustainability (p. 3). She conceptualized scaling in education as four interrelated dimensions: (a) depth, (b) sustainability, (c) spread, and (d) shift of ownership. *Depth* is translated at the teacher level that alters a teachers' beliefs about how students learn and the nature of the subject matter, interactions with students, and underlying pedagogical principles and enacted curriculum, or how teachers and students engage with materials and activities over time (This dimension has been addressed in MSSELL and will be validated in LISTO.). *Sustainability*, the second element of scale, requires that schools move beyond initial implementation to sustain efforts over time (LISTO will reduce efforts from the research/implementation team so that schools can maintain or sustain the work as the i3 funding dissipates. The mechanisms will be in place for their ability to sustain the teacher efforts of the innovations.). *Spread*, the next aspect of scaling, is related to not only the expansion to other classrooms, but also how the innovation principles and norms influence school and district instructional policy and procedures and teacher practices (This element is addressed by PD provided to principals and district instructional leaders, as well as with close communications with the superintendent. Policy and procedural changes will be analyzed over the scope of the grant.). *Shift of Ownership* is the last element in which ownership of the innovation shifts from external to internal reformation. In this phase, school personnel are left with the capacity to sustain, spread, and deepen the innovation principles. Scaling phases move through a 5-year life cycle starting with initial implementation or deepening of the innovation

(Year 1-2), moving to sustaining and spreading (Year 3-4), and spreading and shifting ownership (Year 5). Across the grant cycle, on-site technical support provided to schools is systematically decreased. Initially the original research team will continue to be responsible for PD and on-going support to campus personnel, and will continue to monitor the quality and fidelity of implementation and to collect student outcome data. When PD and support have been owned by school personnel, the innovation will be considered to have been institutionalized. If the educational practices remain effective, widespread diffusion and spread is possible. We expect LISTO will be sustainable beyond the 5-year scope due to the fact that all curriculum, implementation manuals, materials, and MOOPILs via LISTO-Virsity will be in place with easy access by school personnel in the state and nation and will be advertised via the CRDLLA/ ELRC/Aggie STEM websites and placed on Texas Digital Library (open access).

B.1. Unmet Demand for Process, Products, Strategies, & Practices to Reach Level of Scale

The current unmet demands of partner LEAs are noted from a recent needs assessment survey (Appendix J.13) related to LISTO's underlying products, processes, and strategies. A representative sample of the confirmed, participating LEAs, indicated an unmet demand to provide assistance for (a) teacher professional development and improving science achievement, (b) academic language in science, (c) reading comprehension, (d) technology integration into curriculum and instruction, (e) students' higher level thinking and problem solving skills, and (f) hands-on engaged science education for EC, ELL, and mainstream students. Additionally, there were comments that indicated an ongoing demand for support in educating inclusion students, ECs, and ELLs, and ways to decrease the education gap. LISTO meets *unmet demands and offers scale-up opportunities* related to the following ***processes*****:** (a) a *large-scale* longitudinal RCT validation and scaled study will be implemented at rural and non-rural schools that implements the innovation of LIS to determine the impact on science, reading, and writing achievement among EC students, including ELLs, mainstream students, and inclusion students with disabilities; (b) VPD implemented in a *large-scale* longitudinal RCT validation and scaled study with middle school teachers over 5 years that provides professional learning regarding LIS and for principals to lead change in teacher pedagogy for the noted population and a scaled open access MOOPIL platform with LISTO-Virsity accessible to all teachers. LISTO provides the following ***products***: (a) virtual observation instruments for teachers; (b) virtual mentoring and coaching [VMC] guidelines (posted to LISTO-Virsity), (c) APEXIS and Hoot hardware and software platforms, (d) a literacy-infused science (LIS) curriculum provided to teachers that is standards-aligned with technology in instruction with the 5E model, (e) FIS packets with an at-home observation tool and rubric using point-of-view camera glasses, (f) EduSmart Science [TIELLAS with higher-level questions written for the content]; and (g) MOOPILs — LISTO-Virsity with VPD. LISTO provides the following ***strategies and practices*****:** (a) virtual connection of science majors at the university with middle school students and teachers in the classrooms; (b) techniques for improving rural education for middle school students who are EC and/or ELL and inclusion; (c) how to build instructional capacity with teachers; (d) how principals and district instructional leaders can lead, deepen, sustain, and spread the innovation. LISTO, upon validation and if determined to be scalable, has the potential to meet such unmet demands at a national level.

B.2. Addresses Prior Barriers to Reaching Level of Scale

Though LISTO appears to be large and ambitious in scope, the PIs can implement such a project. They have experience in managing large, complex projects, validating a large project, and producing and disseminating results. They have been successful with a validation grant, ELLA-V, across Texas for the past 4 years. They have been faithful in discharging their duties and hiring effectively. The following experiences with barriers (Table 1) were recently shared with i3 regarding scaling. In LISTO, technology will be less of an issue due to APEXIS hardware.

Table 1. *Barriers and Solutions by the PIs*

Prior Experience: Scaling Barriers with ELLA-V	**Solution**
Quick turnaround needed to get positions hired, recruitment of districts/campuses/teachers, obtain districts' memorandum of understanding.	Have personnel, documentation, processes in place by start of school year for the intervention to begin.
Geographic spread (72 campus sites across Texas in eight different school districts in ELLA-V), working out logistics for consent form collection from 61 sites across the state, maintaining confidentiality, Institutional Review Board protocol, and timing to develop testing roster.	Cultivate a vendor relationship with a courier; develop onboarding, testing, curriculum, training, observations in-person regional teacher orientations, virtual orientation materials, and clearly outlined teacher expectations.
The hiring of testers with background checks necessary to work with students.	Obtain districts' substitute list — already have background clearance.
Setting up district network/firewall access so we could access cameras to conduct virtual observations	Engage local district/campus technology specialist.
Some teachers were unfamiliar with the technologies and cameras (cameras on, mics charged and paired)	Clear communication, training videos.
The observation camera company, thereNow, a partner, closed its doors during the third year of the project.	Worked with TAMU legal office and tech specs to keep all platforms and data secure and running.

The research team has disseminated information from the original MSSELL RCT as well as from the ELLA-V grant. There have been five refereed MSSELL publications (Appendix J. 22.), 3 under review, 1 in preparation, and 2 dissertations, and 20 presentations; 15 ELLA (an IES RCT; R305P030032 upon which ELLA-V was based) and ELLA-V publications and 30 presentations. The PIs have experience in facilitating processes for teacher reflection, mentoring, and improvement (one member is author of the Reflection Cycle from *The Principal Portfolio & The Career Advancement Portfolio* and editor for the *Mentoring and Tutoring Journal*, Taylor & Francis) and teacher observation (one member is developer of the virtual observation instrument to be used in LISTO specifically for high-needs students). The PIs also have experience working with external evaluation teams. They have utilized an Advisory Board (AB) with university science, language, and leadership faculty and district personnel to reach a level of scale with validation. The LISTO SRM^2 component will be supported by TAMU with its commitment to high-impact practices for undergraduate students, through “Student engagement in co-curricular activities provides the opportunity to discover the relevance of learning through “real-world application and practice” (TAMU, 2016, ¶10).

B.3. Successful Replication in a Variety of Settings and with a Variety of Populations

The components of the LISTO project will be replicable for variety of school types across Texas at a very reasonable cost. The total cost of such a program is minimal over the 5 years — including a variety of school personnel and diverse parents: 560 teachers, 11,200 students, at least one parent, 140 principals and 40 district instructional leaders (23,140 school participants over time) — approximately $518 per person for implementation. (This is the proximal cost; for the distal cost with teacher rotations of classes they teach beyond the LISTO selected classrooms and students, the number increases to 56,000 students, which makes the distal cost approximately $214 per participant.) Much of the components on “how to” implement such a program on the campus will be available online via the MOOPIL, LISTO-Virsity (Appendix J.12) [The PIs will work with the distribution librarian at TAMU to make the most of traffic on the internet for search engines to locate the materials.]. Included in the distribution for replication across rural, urban, and suburban schools are the products mentioned in B.2. The centers will offer training for teachers in the U.S. with the hardware and software packages related to LIS. Also, school personnel that are in LISTO can actually train others via the MOOPIL concept and that would be the least costly — only the software package would be needed and the strategy of how to develop a MOOPIL (which will be added to our MOOPIL website). All LISTO components of the project will be available for free via LISTO-Virsity and the centers for districts and universities to know how to implement such a program in rural/non-rural settings and with a variety of types of students as indicated.

C. Quality of Project Design/Management Plan: C.1. Clear Goals, Objectives, & Outcomes

The ***GOAL*** of LISTO is to validate, via a longitudinal RCT study, LIS instructional and curricular innovations from a prior research grant with moderate evidence in order to increase instructional capacity of teachers and to improve students' science and reading/writing literacy achievement in rural/non-rural schools for ECs and ELLs. Objectives for LISTO are clear, aligned, and measurable as shown in Table 2.

Table 2. *Objectives, Strategies, Outcomes, and Measures*

Objective 1. To examine the impact and efficacy of bi-monthly LIS VPD for Treatment (T) 5th-GR. teachers compared to routine school district PD for Control (C) teachers on (a) pedagogical skills observed from a low-inference observation tool; (b) lesson effectiveness measured by a fidelity instrument; (c) students' achievement measured by the statewide assessment on reading, writing, and science, standardized and researcher-developed assessments in science; (d) students interest as measured by a researcher-developed Science and Technology Interest Survey [STIS]; (d) qualitative data on the VPD, and (e) qualitative data on VMC from mentor/coach perspectives and mentee/teacher perspectives on quality of coaching and influence on practice. (LIS = literacy-infused science; Hoot Education = Hoot)		
STRATEGIES	***OUTCOMES***	***MEASURES***
Strategy 1.0 Implement VPD bi-monthly for two full years with the same 70 T 5th-GR. teachers; then monthly in the third year; and twice a year in the fourth year (gradually decreasing the VPD for analyzing the sustainability of instructional capacity for LIS).	***Scaled model validated for VPD for 5th-GR. science teachers for the sustainability of instructional capacity for literacy-infused science (LIS).***	**Measure 1.0.** For assessing the impact and efficacy of VPD: (a) observe 3 times per year with the Transitional Bilingual Observation Protocol (TBOP; Appendix J.14-J.17) with 70 T and 70 C teachers; (b) 3 times per year with Science Teacher Observation Record (STOR; Appendix J.18) with 70 T and 70 C; (c) with discourse analysis rubrics for observing the recorded VPD of T teachers; (d) with 10 semi-structured focus groups of T teachers pre/post VPD annually; and (e) with a document analysis rubric with T teachers' virtual portfolio (VFolios via *TK20*; Appendix J.19).

Strategy 1.1. Use Citrix GoTo Meeting and APEXIS technology for VPD (train teachers and testers in use of the soft/hardware solutions).	***Teachers and testers' ease of use with VPD.***	***Measure 1.1.*** Assess the software and hardware ease of use via a short Qualtrics VPD-Use Survey of the T teachers, as well as testers.
Strategy 1. 2. Using Hoot and APEXIS, coordinators observe T and C using STOR for fidelity 3 times (one initial, mid, and ending 28-week period) also used for fidelity of treatment and consequently implement a virtual mentor/coach (VMC bug-in-the-ear program; train T teachers & coaches on Hoot and APEXIS).	***Fidelity of innovation is established, software system has ease of use, VMC model is validated.***	***Measure 1.2.*** There are three measures for this strategy: (a) analyze the fidelity of the innovations using STOR; (b) assess the software and hardware ease of use via a short Qualtrics VMC-Use Survey of the T teachers, and (c) with an expert coach, assess, via VMC rubrics, the mentor/coaches' real-time feedback sessions with the T teachers and perceived influence of VMC.
Strategy 1.3. Provide a monthly training on building instructional capacity with VPD using Citrix and APEXIS technology for principals.	***VPD model for principals of 5th-GR. science teachers for the sustainability of instructional capacity for LIS.***	***Measure 1.3.*** VPD for principals is measured by: (a) assess the software and hardware ease of use via a short Qualtrics VPD-Use Survey of the T principals and (b) with 10 semi-structured focus groups of T principals pre and post VPD annually.
Strategy 1.4. Develop teachers as reflective practitioners (teach them the Reflection Cycle [Irby & Brown, 2001]).	***Teachers who can improve via reflection on their practice.***	***Measure 1.4.*** Data from reflections in portfolios will be gathered in TK20 (portfolio/data management) and analyzed qualitatively.
Strategy 1.5. Determine the sustainability of VPD of 5th-GR. T students matched annually across 4 cohorts by school characteristics.	***A curriculum innovation in LIS for rural/non-rural high needs students.***	***Measure 1.5.*** STAAR science, reading, writing, ITBS Science, Big Idea Science Assessment (BISA) pre/post, science benchmark tests, and writing rubrics at 5th GR.
Strategy 1.6. Determine differences in T students by student and school characteristics annually.	***A curriculum innovation in LIS for rural/non-rural high needs students.***	***Measure 1.6.*** STAAR science, reading, writing, ITBS Science, BISA pre-post, science benchmark tests, and writing rubrics at 5th GR.
Strategy 1.7. Determine differences between T and C students on science interest and achievement.	***A STIS instrument developed for use.***	***Measure 1.7.*** STIS, STAAR science, reading, writing, ITBS Science, BISA, pre/post, science benchmark tests, and writing rubrics at 5th GR.
Objective 2. Determine the academic sustainability of LIS that occurred with the first cohort of T 5th-GR. students over 4 years when they matriculate to 6th, 7th, and 8th GRs.		
STRATEGIES	***OUTCOMES***	***MEASURES***
Strategy 2.0. Implement VPD monthly with the teachers of the first cohort of the 5th GR. as they	***Scaled model for minimal VPD with 6th, 7th, and***	***Measure 2.0.*** The following measures will collectively assess the efficacy of the VPD: (a) 3 times per year with

move to 6th, 7th, and 8th GRs. for analyzing the sustainability of instructional capacity for LIS and assess the impact of sustainability on the participating students across time by grade level and overall by ending GR. 8 and by school characteristics.	***8th-GR. teachers developing instructional capacity for teachers to maintain the curriculum of LIS for students who are EC and ELL in rural and non-rural schools.***	STOR with 70 T and 70 C teachers; (b) with discourse analysis rubrics for observing the recorded VPD of T and C teachers; (c) with 20 semi-structured focus groups of 140 T and C teachers pre and post VPD annually; and (d) compare the T and C students' achievement measured by STAAR science (8th-GR. level), STAAR reading and writing (GRs. 6-8), and ITBS Science (GRs. 6-8).
Strategy 2.1. Use GoToMeeting and APEXIS technology for VPD (train teachers by 6th-, 7th-, and 8th-GR. levels with Cohort 1 with use of the software/hardware).	***Scaled model for VPD with GoToMeeting and APEXIS.***	***Measure 2.1.*** Assess the software and hardware ease of use via Qualtrics VPD-Use Survey of the T and C teachers.
Strategy 2.2. Using the APEXIS technology, coordinators observe T and C using STOR for fidelity 3 times and provide feedback to teachers on their level of implementation of LIS.	***A fidelity instrument for increasing instructional capacity of science teachers.***	***Measure 2.2.*** There are two measures for this strategy: (a) analyze the fidelity of the innovations using STOR and (b) assess the software and hardware ease of use via a Qualtrics VMC-User Survey of the T and C teachers.
Strategy 2.3. Provide 45 hours of VPD using Citrix and APEXIS technology for the same school principals over 3 years for 6th, 7th, and 8th GRs. targeted to build the teachers' instructional capacity specifically for EC and ELL students, in LIS.	***A leadership model for principals to increase instructional capacity for teachers in science.***	***Measure 2.3.*** The VPD for principals will be measured in the following ways: (a) assess the software and hardware ease of use via Qualtrics VPD-Use Survey of the T and C principals and (b) with 10 semi-structured focus groups of 70 T and C principals pre and post VPD annually.
Objective 3. Provide scaled outcomes with dissemination of products, training, and sustainability beyond the conclusion of Project LISTO.		
STRATEGY	***OUTCOMES***	***MEASURES***
Strategy 3.0. Foster via Hoot and blogs, a community of practice dyad (match dyads of teachers across rural, urban, and suburban campuses) for improving practice in LIS.	***Model of peer mentoring and professional learning dyads.***	***Measure 3.0.*** Take data from Hoot and blogs. Use a researcher-designed rubric for video and document analysis. Use research-developed teacher survey related to dyads of practice — used with NVivo.

Strategy 3.1. Develop LISTO-Virsity (based on the work from ELLA-V) as an outgrowth for dissemination and scaling and provide all webinars in a venue called a MOOPIL; LISTO-Virsity (which will house webinars from T teachers, coordinators, mentor/coaches, and principals). LISTO curriculum will be posted each year under CRDLLA/ELRC/Aggie STEM.	***MOOPIL site developed and advertised statewide with teachers in the project making presentations on the project components; > the basic 30hr/yr. of Continuing Ed hours we provide in LISTO VPD.***	**Measure 3.1.** Survey teachers who are visiting the MOOPIL site, LISTO-Virsity for ease, use, and relevance for classroom practice. LISTO-Virsity (virtual targeted PD university) housing multiple webinars for free for the 560 middle school science teachers across Texas deduced from number of students in GRs. 5, 6, 7, and 8). Assess the number of hits on the LISTO-Virsity webinars from CRDLLA, ELRC, and Aggie STEM websites.
Objective 4. To determine the influence of the components of FIS and SRM[2] for families,		
STRATEGIES	***OUTCOMES***	***MEASURES***
Strategy 4.0. Using FIS, engage family members and respective T students at 5th-GR. in science dialogue related to (a) use of academic language, (b) use of misconceptions in science, (c) level of engagement, (d) increase in student's vocabulary, (e) level of satisfaction with FIS, and (f) attitude toward science with FIS.	***FIS packages available for school use (open access) with suggestions for point-of-view camera glasses use.***	**Measure 4.0.** Observation rubric developed to assess academic language use, misconception discussions in science, level of engagement, and student-level vocabulary improvements over time; video analysis input in NVivo; semi-structured survey on satisfaction and attitude from parent and child on FIS (pre and post component survey — Likert scale)
Strategy 4.1. School/university partnership that (a) engages and motivates students to become scientists (T students) based on intervention from the university science majors, and (b) T students' and university science majors' perspectives of the SRM[2].	***A model of school/ university partnership that increases engagement in science and motivation to become a scientist.***	**Measure 4.1.** Observation rubric per unit of school/university partnerships with university science majors engaging with middle school students; interviews with middle school science students, their teachers, and the university science majors on their perception of positive and negative aspects of SRM[2].

C.2. Adequacy of Management Plan to Achieve Objectives on Time and Within Budget

There are four management groups involved in LISTO. The *Advisory Board (AB)* ensures application in the schools for furthering the project and disseminating findings. The PIs serve as the *Policy and Procedural Oversight Group (PPO)* and will do all ordering of equipment, preparing contracts, MOUs, and intervention materials with the College and TAMU Post-Award

Research Office, and they will train coordinators, work with the evaluation team in evaluation components, and will hire the implementation personnel. PIs will not be involved in data analysis as that will be supervised by the evaluation team; any data coming in to CRDLLA, ELRC, Aggie STEM will be monitored carefully by an evaluation specialist at TAMU, but hired by JHU for oversight. The *Application/ Implementation Group (AIG)* is made up of research associates/mentors/coaches, graduate research assistants, undergraduate research assistants, technical support at TAMU, and consultants, and they will deploy all interventions with project teachers and principals and utilize all technical support and consultants to implement VPD, VMC, and VOBS. The *Evaluation Group (EG)* consists of the external evaluator team and its internal evaluators and graduate assistants, and will be chaired by JHU staff. This group is charged with gathering data and implementing the evaluation design and analysis described in this application. Table 3 includes the major management milestones with objectives and strategies, group responsible, and timeframe.

Table 3. *Major Milestones, Group Responsible, and Timeframe*

Major Milestones	**Responsible**	**Yr 1**	**Yr 2**	**Yr 3**	**Yr 4**	**Yr 5**
Objective 1.0						
Strategy 1.0	**PPO;AIG;EG**	√	√	√	√	√
Strategy 1.1	**PPO;AIG**	√	√	√	√	
Strategy 1.2	**PPO**	√	√	√	√	
Strategy 1.3	**AIG; EG**	√	√	√	√	
Strategy 1.4	**EG**	√	√	√	√	√
Strategy 1.5	**AIG**	√	√	√	√	
Strategy 1.6	**EG**		√	√	√	√
Strategy 1.7	**EG**		√	√	√	√
Objective 2.0						
Strategy 2.0	**EG;AIG**	√	√	√	√	√
Strategy 2.1	**EG; AIG**	√	√	√	√	√
Strategy 2.2	**AIG;EG**	√	√	√	√	√
Strategy 2.3	**AIG;EG**	√	√	√	√	√
Objective 3.0						
Strategy 3.0	**EG; AIG**	√	√	√	√	

Strategy 3.1	**AIG**	√	√	√	√	√
Objective 4.0						
Strategy 4.0	**EG; PPO**	√	√	√	√	√
Strategy 4.1	**AIG**	√	√	√	√	√
Other Critical Components						
Recruit/hire all personnel	**PPO**	√				
Establish Advisory Board	**PPO**	√				
Establish all subcontracts/MOUs	**PPO**	√				
Order all materials	**PPO**	√	√	√	√	√
Establish all training with specific vendor and partners	**PPO**	√	√	√	√	√
Communicate with district/ school administrators	**PPO**	√	√	√	√	√
Establish final agreements with districts/schools	**PPO**	√				
Work with i3 Project Officer on Mgt. Excel File	**PPO**	√				
Meet with Project Office monthly	**PPO;AIG**	√	√	√	√	√
Meet with EG four times annually	**PPO;EG**	√	√	√	√	√
All data collection/ analysis/reporting	**EG**	√	√	√	√	√
Grant reporting	**AB; PPO;EG**		√	√	√	√
Disseminate results	**PPO;EG;AIG; AB**	√	√	√	√	√

AB-Advisory Board; AIG-Application/Implementation Group; EG-Evaluation Group; PPO-Policy/ Procedural Oversight Group

C.3. Clarity and Coherence of Financial and Operating Model and Plan

At the end of LISTO, the total number of students who will have engaged in the project is 11,200, the total number of teachers will be 560, and the total number of principals will be 140. Approximately 40% of the total budget will directly support PD and teacher mentoring/coaching (Objective 1), with the remainder pointed toward the intervention, testing, and evaluation of LISTO goals (Objective 1). The components of the LISTO project will be able to be replicated by the 3 centers from TAMU for schools across the country at a very reasonable cost. The total cost of such a program is minimal, including teachers, students, at least one parent, principals, district instructional leaders (23,140 school participants over time — approximately $518 per person for implementation). This is the proximal cost; for the distal cost with teacher rotations of classes they teach beyond the LISTO selected classrooms and students, the number increase to 56,000 students, which makes the distal cost approximately $214 per participant.

LISTO has far-reaching cost-related impacts with innovations that have the potential to expand to a total of over 1.1 million Texas middle school students in GRs. 5-8, particularly to the rural schools with EC students and ELLs (TEA, 2016). The MOOPILs, via LISTO-Virsity, will be provided through the three Centers and has the potential to reach over 300 middle school science teachers in the local TAMU area alone. Across Texas, in GRs. 5-8, there are approximate 2000 middle school science teachers (extrapolated from Texas Academic Performance Reports [TAPR]; TEA, 2016). There are approximately 211,000 science teachers nationally per NCES (2011). Indirect impact increases with anticipated large scale-up costs of less than $11 per student and teacher across Texas only.

C.4. Ensuring Feedback and Continuous Improvement in the Project

LISTO PIs and personnel have strong experience running large research grants, as well as training projects at the federal and state levels. Continuous improvement has come in the form of communication with district superintendents, curriculum directors, principals, teachers, the evaluation team, and the local project team. Every 2 weeks, the team will be in communication with the treatment teachers in LISTO. Every semester, there will be testers on the campuses, and the PIs and/or personnel will meet via GoToMeeting with every superintendent or principal in the participating schools. We will continue that communication process for the entire life of LISTO. The research team will meet monthly to discuss updates and improve the project; a Milestone and Management Chart will be kept in the PIs' office. Additionally, we will provide written notes and an agenda on our monthly meetings with the i3 officer. All materials will be delivered in a timely manner. We have been faithful in the current i3 validation project (ELLA-V) and have presented at i3 meetings in Washington, D.C. over the past 4 years and would continue to do so for LISTO. We understand the importance of continuous improvement for such a project in order to improve processes and products.

D. Quality of the Project Evaluation: D.1. Project Will Produce Evidence of Effectiveness

Sample and Research Design — The overall LISTO project is a quantitative-dominant, mixed-methods research project and is symbolized as concurrent, QUAN+qual research design (Johnson & Onwuegbuzie, 2007) in which there is less emphasis on how one strand informs the other; rather, the focus is on interpretation of conclusions from both, or concurrent, strands. To evaluate the effectiveness of the intervention, we will use a clustered randomized design where schools will be randomly assigned to either treatment or control condition. We plan to recruit 70 Texas schools in rural and non-rural (urban/suburban) settings. We identified districts using the federal guidelines for RLIS program to identify eligible districts in Texas and contacted these districts about project participation. EC students in these rural districts make up 50% to 93% of the population, and ELL students are 8.7% to 49%. Urban/suburban districts were contacted based on the fact that they serve a high percentage of EC students (42% to 99.7%) and ELL students (6% to 68%) (Appendix J.2). Sixty-six percent of the students to be served will be enrolled in rural schools within the participating rural districts so as to meet *Absolute Priority 4*. Therefore, we propose to include 46 rural schools all randomly assigned (by the EG) to treatment (T=23) or control/comparison (C=23) conditions; and 24 non-rural urban/suburban schools all randomly assigned (by the EG) to treatment (T=12) or control/comparison (C=12) conditions. The integrity of such assignment will be maintained because when a school is assigned to receive T in GR. 5 in Year 1, then this school will continue to receive T in the subsequent years; the same is true for C schools. An average of two GR. 5 teachers per school will be randomly selected by the EG at the randomized campus to implement the corresponding condition. In this manner, there will be no contamination of both conditions on the same campus; therefore, we address the issues of a design flaw noted by Song and Herman (2010) by separating out the intervention from the teacher effects since the teachers will not be involved in both conditions. School-, teacher-, and student-level demographic data will be collected including: ethnicity (teacher/student/principal), gender (teacher/student/principal), socioeconomic status (school/student), regional classification (e.g., rural, urban, suburban), language spoken at home (student), certification (teacher), demographics of teachers, and inclusion student or not. All the measures will be administered in the same way to all participants each year GRs. 5-8, T and C. To test the **sustainability** of this project, we will have two tiers. ***Tier 1*** is a longitudinal design following the same 5th-GR. teachers – 140 teachers (2 teachers per each of 70 schools) will be included with four GR. 5 cohorts over 4 academic years with a total of 11,200 students in GR. 5 (140 teachers x 20 students per class = 2,800 students x 4 years). Such sustained impact will be examined at student level as each year a new cohort of 5th-GR. students will join the project and will be matched on student-level characteristics and their baseline pre-tests, and they will be compared on science, reading, and writing achievement as a result of their teachers' pedagogy and the curriculum intervention. ***Tier 2*** is a longitudinal design following the same GR.5 students over 4 years. Cohort 1 of 2,800 GR. 5 students who participated in the first year of implementation will be followed as they matriculate through GRs. 6, 7, and 8. Across those 3 years for Cohort 1 participants, their treatment (T) and control (C) teachers (anticipated 140 additional teachers per year) will receive VPD and VMC innovation on how to infuse literacy into science, along with science content. In this way, teacher effect in GRs. 6-8 between T and C can be accounted for, and the only difference, if any, at student level by the end of GR. 8 would be due to the sustained impact of the full intervention that occurred in GR. 5 in the T schools. As participating students enter and leave the schools, especially if they transfer among schools, we will keep careful track of their initial assignments and their receipt of services over time. Participating students at all T and C schools will be pre-tested at the beginning of Project LISTO as baseline, and at the end of each school year of intervention (GRs. 5-8), respectively.

Attrition and Missing Data — In our MSSELL NSF project, because it was a pure longitudinal study with no cohorts added after the initial one was established, the overall attrition rate was around 35%, with an average overall attrition rate of 17% per year over 2 years, and a differential attrition rate of 3% between T and C groups at the student level, 0% attrition rate at the school level, and a 2% attrition rate of teachers. Such a combination of overall and differential attrition rate falls within low levels of expected bias outlined by the What Works Clearinghouse Procedures and Standards Handbook (2014). Further, attrition in the original MSSELL was more likely related to random causes such as family mobility or school schedule of the intact cohorts at the schools during the 2 years of the pure longitudinal study. We also examined the attrition rate from our current i3 validation study (which is ending in 2017), and we have been able to maintain an overall cluster attrition rate of less than 3% and a differential cluster attrition rate of less than 1% over the 3-year period. Therefore, we do not anticipate a high overall attrition rate in LISTO. Further, the EG will follow the model proposed by WWC (2014) to determine the overall and differential attrition that may bias the estimated intervention effect. EG will use an intent to treat model, following all participating students in all schools randomly assigned at the outset. Finally, to meet WWC (2014) evidence standards, we will not impute missing outcome data in any analyses. The analysis sample is defined as all cases with non-missing outcome data.

Fidelity of the Implementation of the Intervention — STOR, developed and tested in MSSELL, is an instrument used to monitor the fidelity of the intervention with specific observational questions appropriate for C schools as well. STOR has internal consistency of .94. It captures the core components of intervention three times per year. In addition, virtual VOBS using TBOP (Lara-Alecio & Parker, 1994) will be conducted 3 times annually in both T and C classrooms also as part of the fidelity of the implementation. The four TBOP observational domains are Activity Structure, Communication Mode, Language Content, and Language of Instruction (Lara-Alecio, Tong, Irby, & Mathes, 2009). Data collected via TBOP or STOR will be incorporated as teacher-level characteristics into the impact analysis.

D. 2 and D.4. Key Questions and How Each is Addressed; Clear Credible Analysis Plan

Power Analysis — In Project LISTO, because students will be nested in classrooms, which in turn are nested within schools, we used Optimal Design (Raudenbush, Liu, Spybrook, Martinez, & Congdon, 2006) to determine the number of schools needed for the experiment. The ability to detect a treatment effect at a certain level of power in a hierarchical linear modeling (HLM) framework depends on several factors: intra-class correlation (ICC, ρ), the correlation between pre and posttests (r), and the average number of students in each school (n). In our power analysis, the parameters included an α=.05, pre-posttest correlation of .70, a target minimum detectable effect size of .25 (although the effect sizes in our experimental research have mostly been between .35 to .7, we decided to use a more conservative effect size in the validation study), a cluster size of 40 (including two teachers/classes per campus), and ICC of .10, which is quite commonly found in cross-sectional studies such as the Special Education Elementary

Longitudinal Study (see Hedges & Hedberg, 2007). Using these parameters and taking into consideration the potential attrition rate over time, by the end of the project year, we can detect an effect size of 0.25 with a power of over 0.80 if we start with a school sample of 35 (23 rural and 12 urban/suburban) per condition or 70 overall.

Data Collection — Data collection by the EG will occur in Years 1-5. At the teacher level, VOBS and field notes will be conducted three times (beginning, mid, and end) annually in all classrooms. Other data including principal and teacher interviews and surveys, and reflections in teacher portfolios will be collected annually. During Year 1 of the project, the qualitative component of the project will be refined including measures, trustworthiness, and credibility of the data. Student scores on standardized assessments such as ITBS, as well as BISA will be collected at the beginning and end of GR. 5 each year in Years 2-5; scores on benchmark tests will be collected every nine weeks in GR. 5 of each year in Years 2-5. Student scores on state-mandated assessments such as STAAR science, reading, and writing will be collected in the spring of GR. 5 each year in Years 2-5. Further, for the first cohort of students, STAAR reading and writing will be collected in the spring of GRs. 6-8; STAAR science will be collected in GR. 8; and scores on ITBS science will be collected annually. LISTO will address specific research questions aligned with the project objectives and strategies. There are two major confirmatory questions (Appendix J.20). In the following section, Strategy = S.

Questions to Evaluate Objective 1 and the Data Analysis per Question—<u>1A.(S1.0).</u> To what extent does 5th-GR. science teachers' instructional delivery differ between T and C classrooms as measured by TBOP annually? *<u>Analysis:</u> to answer this **confirmatory question**, chi-square test of homogeneity of proportion to identify if the proportion of each category under every domain in TBOP is homogenous between T and C classrooms*. **<u>1B.(S 1.0)</u>** To what extent do teachers' instructional practices improve annually as a result of VPD as measured by STOR? *Analysis: A paired-sample t-test to identify the improvement on STOR annually for T teachers*. **1C.(S 1.0, 1.1., 1.4)** What is the teachers' perceived effectiveness of the VPD, and based on the VPD, do they perceive their practice to improve with reflections included in training? *Analysis: Phenomenological study (Creswell, 2014) with data, researcher, and methods triangulation and low-inference descriptors (Burke-Johnson, 1997) reported to address credibility (internal validity). Data collected via field notes, classroom observations, semi-structured, open-ended surveys and interviews (Lincoln & Guba, 1985) and/or teacher reflections/portfolios (via TK20 portfolio collection software). Data analyzed using constant comparative method (Cresswell, 2014) and coded according to themes for identifying trends or patterns with all data entered into NVivo software. Focus groups with T teachers at GR. 5 will be conducted; patterns will be drawn, description of the relationships both formal and informal will be conducted, meanings both tacit and explicit will be sought, and the ability to implement and sustain such interventions within other schools will be analyzed.* **1D.(S 1.2)** To what extent do teachers implement the innovations with fidelity? *Analysis: Descriptive statistics.* **1E.(S 1.2)** How do teachers perceive the ease of use and quality of VMC using the Hoot Education and APEXIS software and hardware? *Analysis: (See 1C).* **1F.(S 1.3)** What is principals' perceived effectiveness of the following components: VPD, interventions, observation tools, teacher implementation, and student achievement? *Analysis: (See 1C with exception of teacher portfolios- not examined for principals).* **1G.(S 1.3)** What is the principals' perceived effectiveness of their ability to build teacher capacity based on training and does their capacity increase annually? *Analysis: (See 1F).* **1H. (S 1.7)** To what extent do students differ between T and C on science, reading/writing achievement and science interest in after one year of intervention GR. 5? *Analysis: To answer* *this* ***confirmatory question****, we will use hierarchical linear model (HLM) to analyze the treatment effects in GR. 5 after year 1. Student will be the level-1 unit of analysis, with pre-test score as covariate (ITBS and STIS); teacher will be the level-2, and school as level-3 unit of analysis. The condition of T or C will be included as level-3 predictor as school is the unit of randomization. A simple presentation of the model follows:* $Y_{ijk} = \gamma_{000} + \gamma_{100}Pretest_{ijk} + \gamma_{001}Treatment_k + r_{0jk} + u_{00k} + e_{ijk}$, *STAAR science is first administered in the spring of Gr. 5, so ITBS pre-test will be used as covariate (this will be exploratory as ITBS is used to measure the domain of science); for STAAR reading and writing, scores from STAAR Gr. 4 will be collected and used as covariates.* **1I.(S 1.6)** Do student and school characteristics predict T students' science and reading/writing literacy achievement annually? *Analysis: HLM will be used with student level (ELL/EC/mainstream/inclusion) and school level characteristics (rural/non-rural) included as independent variables, pre-test as student-level covariate, and post-test as outcome variable.* **1J. (S 1.5, 1.7)** Does the impact of LISTO become stronger as T GR. 5 teachers continue exposure to intervention across the years as reflected on rural and non-rural students' science, reading/writing literacy achievement and science interest? *Analysis: because there will be student joiners each year (taught by the same Gr. 5 teachers), student-level characteristics will be matched and then the matched sample will be compared using HLM with Gr. 5 pre-test score as covariate and Gr. 5 post-test score as dependent variable, cohort (e.g., Year 2, Year 3, etc.) as student-level variable, and rural/non-rural as school-level variable.* **1K.(S 1.1)** How do testers perceive the effectiveness of virtual training on implementing the testing of ITBS, BISA, and Benchmark Assessments? *Analysis: (See 1C).*

Questions to Evaluate Objective 2 and the Data Analysis Per Question — 2A.(S 2.0) To what extent does the first cohort of students differ between T and C conditions on science, reading, and writing achievement when they matriculate into GR. 6, GR. 7, and GR. 8 respectively (by student and school characteristics)? *Analysis: Because the first cohort of students is taught by different teachers at each GR. level, i.e., GR. 5, GR. 6, GR. 7, and GR. 8, the data are cross-classified in nature. Although traditional HLM and cross-classified random effect models effectively handle data dependencies due to student and classroom effects, they make the unrealistic assumption that there are no persistent or constant classroom effects across time. Such assumptions can be relaxed in the recent development of n-Level structural equation modeling (n-Level SEM; Mehta, 2013a). Therefore, we will estimate a 4-level linear growth model, with the levels corresponding to student and grades using the R package xxm (Mehta, 2013b), which implements model estimation under the n-Level SEM framework. Such model can be translated into the more commonly used multilevel notation as*

$$y_{i(j_1,j_2,j_3)t} = \alpha_{000} + \alpha_{100} TIME_t + \gamma_{11} Treat_{(j_1,j_2,j_3)} + \gamma_{21} TIME_t \times Treat_{(j_1,j_2,j_3)} + e^{(t)}_{i(j_1,j_2,j_3)t} + u^{(s)}_{0i(j_1,j_2,j_3)} + u^{(s)}_{1i(j_1,j_2,j_3)} TIME_t + \sum \lambda_{tl} \eta^{(l)}_{1j_l},$$

We compared different types of mixed models (including traditional HLM, CCREM, and the n-Level SEM) using data collected from a previous longitudinal RCT and have found that the n-Level SEM yields the most accurate estimates (Lai, Tong, Yoon, Lara-Alecio, Irby, & Kwok, 2014). **2B.(S 2.1, 2.2)** What is GR. 6, GR. 7, and GR. 8 teachers' perceived effectiveness of VPD? *Analysis: (See 1C).* **2C.(S 2.0)** Do GR. 6, GR. 7, and GR. 8 teachers' instructional practice improve annually per each respective year as a result of VPD as measured by STOR? *Analysis: Paired-sample t-test to identify the improvement on STOR each year.* **2D.(S 2.3)** Do GR. 6, GR. 7, and GR. 8 principals' perceive their ability to increase instructional capacity annually as a result of their own VPD on LISTO? *Analysis: (See 1C).*

Questions to Evaluate Objective 3 and the Data Analysis Per Question — 3A.(S 3.0) How do teachers perceive the community of practice and the blogs for improving their own practice in

relation to science education? *Analysis:* *(See 1C).* **3B.(S 3.1)** How do teachers who access LISTO-Virsity perceive the usefulness and relevance of LISTO-Virsity? *Analysis: Number of hits on the LISTO-Virsity website, and an evaluation form per LISTO-Virsity PD online so that teachers can obtain time on their continuing ed hours needed by the state for their evaluations.*
Questions to Evaluate Objective 4 and the Data Analysis Per Question — 4A.(S 4.0). In what ways do family members and respective T students at 5th-GR. level engage in science dialogue through FIS related to (a) use of academic language, (b) use of misconceptions in science, (c) level of engagement, and (d) increase in student's vocabulary as measured by video analyses from the parent's and child's perspectives using recordings from point-of-view camera glasses? *Analysis:* *(See 1C with rubrics developed for video analysis of point-of-view glasses recordings).* **4B.(S 4.0)** What is the satisfaction level of family members in terms of working with their children at home on standards-aligned and engaged science activities in FIS, and do students' attitude toward science increase based on the FIS activities? *Analysis:* *(See 1C, with rubrics developed for observing in the video recordings from point-of-view glasses, the attitudes of students and parents toward science).* **4C.(S 4.1)** What are perceptions on science engagement and motivation to become a scientist (the middle school T students) from the (a) university science majors' and (b) the middle school T students perspectives in the SRM^2 component? *Analysis:* *(See 1C, with the exception of teacher portfolios).*

D.3. The Project Evaluation Will Study Scale in Diverse Settings and Diverse Populations

Seventy schools will be selected from rural/urban/suburban districts that serve high-needs students. The independent evaluation will be directed by the LISTO Logic Model (Appendix J.21.). The key elements and approach of MSSELL was in an urban district, *but in LISTO*, there will be a majority of rural school students, together with other urban/suburban schools included.

Also, an altered condition will be that the PD is delivered virtually and not on site which provides evidence for scaling LISTO VPD across schools in multiple locales. Another altered condition is the observation and mentoring of teachers via virtual tools (VMC). Such VPD and VMC have already been implemented in a current i3 validation project with positive teacher feedback and improved teaching practice (Tong, Irby, & Lara-Alecio, 2015; Tong et al., 2016).

D.5. Clear Articulation of Plan and Measurable Threshold of Outcomes

Outcome Measures — We will compare students' achievement on the constructs for these constructs and the psychometric properties of the corresponding tests) aligned with the research questions as measured by both (a) researcher-developed assessment of writing rubrics (developed and validated in MSSELL, Appendix J.7) to measure and compare the academic language development and conceptual understanding through science notebooks; district benchmark test in science; and BISA (developed in MSSELL with satisfactory internal consistency); and interest inventory, STIS; (b) state-mandated standards-aligned assessments, including STAAR science (GR. 5 & GR. 8), STAAR reading/writing (GR. 5-8); (c) rigorous standardized science instrument to measure progress in science, pre- and post-ITBS to GRs. 5-8 students. Among these, benchmark tests and BISA are curriculum-based and formative assessments aligned with state and national standards and are embedded in instruction to provide timely feedback for purposes of adjusting instruction to improve learning. STAAR measures academic progress of all students, inclusive of ELLs and EC students. ITBS science assesses not only students' knowledge of scientific principles and information, but also the methods and processes of scientific inquiry. Most of above-mentioned measures were used in MSSELL. Adopting the multilevel, multifaceted assessments framework on science achievement by Ruiz-Primo, Shavelson, Hamilton, and Klein (2002), we propose two levels of assessments: *proximal* *level* assessment which is used to ensure that the teachers are teaching the assigned standards (and to hold schools accountable), i.e., district benchmark tests, BISA, and science notebooks; and *distal level* assessment based on state or national standards in a particular domain, i.e., STAAR and ITBS. At the teacher level, we will compare instructional effectiveness between T and C teachers using TBOP and STOR. Other qualitative rubrics will be developed.

D.6. Sufficient Resources to Undertake the Project Evaluation

The evaluation will be conducted at JHU, one of the nation's premier research institutions. We have been in conversation with JHU, and they have provided the evaluation and the budget to undertake such a task. The Center for Research and Reform in Education (CRRE) at JHU will provide strong and consistent organizational support. Resources follow: ***Support and Commitment.*** JHU provides multiple levels of support to the CRRE, which employs five Ph.D.s and five other research and support staff engaged in a wide range of research involving children from preschool through high school who are in high-poverty communities. CRRE PIs are full-time researchers without teaching responsibilities who are therefore able to focus on high-quality longitudinal research, including many randomized and matched field experiments. ***Program Assistance.*** JHU offices assist in the effective and efficient operation of the CRRE. JHU's Homewood Research Administration administers grants and contracts; the Controller's Office provides accurate and timely monthly statements for budgetary control; HR assists in hiring and salary administration; Library Services provides a full range of assistance in research and document procurement, and the Office of Communications assists in national dissemination. ***Resources.*** CRRE's office is located off campus, and therefore qualifies for low off-campus overhead rates. JHU's facility provides office administrative and clerical services; photocopying and conference rooms; and areas for maintaining and analyzing data, inventories of center publications, and reference publications. ***Major Evaluation Personnel for Effective Evaluation Assigned to LISTO.*** *Steven Ross, Ph.D*., will lead the randomized control trial of the program. He is currently senior research scientist and professor at the CRRE at JHU, the author of six textbooks and over 125 journal articles, and Editor Emeritus of the research section of the *Educational Technology Research and Development* journal. He was the first recipient of the Eminent Faculty Award and the Lillian and first Morrie Moss Chair of Excellence in Urban Education when at The University of Memphis. He has testified on school restructuring research before the U.S. House of Representatives Subcommittee, has been a consultant to NSF on evaluation design, and is a technical advisor or researcher on current federal and state evaluations. *Alan Cheung, Ph.D*., Professor, CRRE at JHU, also Professor and Dean of School of Education at the Chinese University of Hong Kong, has participated in large-scale national randomized field experiments, including RCTs with ECs and ELLs. He is an expert on data analysis, including HLM, and experimental designs; has worked on evaluations in China, Ireland, and the U.S.; has published over 100 journal articles, book chapters, and technical reports; and was a recipient of the Palmer O. Johnson Award in 2008. He will develop and oversee an overall plan for all aspects of the LISTO evaluation, to ensure integration and quality of the evaluation as a whole. *Rebecca Wolf, Ph.D.*, Assistant Professor, has expertise in educational program evaluation, quantitative research design (including RCT, quasi-experiments, and mixed methods), statistical methods, and data management. At the CRRE, she manages the quantitative design and analyses for two i3 validation studies. She has also conducted evaluations of teacher professional development and teacher preparation programs, principal professional development programs, high school career pathways, charter school networks, and state school finance policy.