Table of Contents

Virtually-Infused Collaborations for Teaching and Learning Opportunities for Rural Youth: Implementation and Evaluation of Online and Face-to-Face Delivery in High-Needs Schools (Project VICTORY)

Texas A&M Research Foundation and the Center for Research & Development in Dual Language & Literacy Acquisition (CRDLLA), College of Education and Human Development (CEHD), Department of Educational Psychology, will collaborate with the Education Leadership Research Center (ELRC) and Aggie STEM and Johns Hopkins University's Center for Research and Reform in Education (CRRE) for ***Virtually-Infused Collaborations for Teaching and Learning Opportunities for Rural Youth: Implementation and Evaluation of Online and Face-to-Face Delivery in High-Needs Schools (Project VICTORY, VP)***. Private partners are LogMeIn, Dr. Janice Koch, CRRE, Inline Resources, Ichor Solutions, and Nearpod (see Appendix G). CRDLLA will also provide a portion of the cost-share match.

On January 20, 2020, the COVID-19 pandemic was just beginning with the first-known U.S. patient testing positive for Coronavirus in Washington state, followed with the first school closure on March 5; by May 11, 48 states had orders or recommendations for school closures, impacting approximately 50.8 million public school students (Peele & Riser-Kositsky, 2020). Virtual teaching, assisted teaching from parents, and learning by students were thrust upon the nation's educators, parents, and students with often a matter of days without preparation. With hands-on science, the situation has been grave since science for elementary children, in its best form, is typically a socially, collaborative, hands-on engaging activity, and in this virtual world, yet, it was reportedly more knowledge and concept learning from discussion and teacher demonstration, as opposed to hands-on engagement. The **overarching goal of VP** is to determine the impact of virtual and face-to-face science teaching and learning in rural elementary schools via a longitudinal, randomized controlled trial study (RCT). VP, a mid-phase project, will serve over its 3 years, 60 campuses from the 418 Texas Rural and Low-Income School program-eligible districts (RLIS; see Appendix C for ***77 rural local education agencies'*** [LEA] partner commitment letters; these districts house 112 elementary/intermediate schools) with cumulatively 9,180 teachers (180), high-needs 3rd, 4th, and 5th grade students (4,500), and families (4,500) by building science and literacy instructional capacity, improving students' science and literacy performance and interests, and increasing family involvement in science. The VP longitudinal model includes students who are economically challenged (EC) students eligible for free or reduced-price meals and/or English learners (ELs). In Texas, 60.6% of students are EC, and 19.4% are EL (TEA, 2020). The proposed RLIS districts serve large numbers of EC students (44% to 100%) and are inclusive of ELs (2% to 74%); districts' characteristics and map are listed in Appendices C.1 and C.2. ***It is worth noting, RCT studies addressing rural STEM education and RCT studies on online vs. face-to-face delivery of science for elementary rural students are almost non-existent,*** particularly with the included high-needs student/family groups of ELs and ECs. VP will fill an important gap in our understanding of rural science education and online vs. face-to-face education, will contribute to evidence-based research for improving, particularly rural high-needs elementary students' engagement and achievement, and should improve teachers' instructional capacity in literacy-infused science and online instructional capacity. CRRE's independent evaluation is expected to report critical evidence about VP's effectiveness via the RCT with the Fidelity of Implementation (FOI) processes developed and published from CRDLLA (Tang et al., 2020;

Tong et al., 2019; 2020). CRRE has designed the evaluation component to meet WWC standards without reservations. ***ABSOLUTE PRIORITIES:*** VP is based on successful research from two prior grants: Project Middle School Science for English Language Learners (MSSELL, NSF, DRL-0822153) and English Language and Literacy Acquisition-Validation (ELLA-V, i3 Office, PR/Award Number U411B120047). Tong, Irby, Lara-Alecio, Guerrero, Fan, and Huerta (2014) and Lara-Alecio, Tong, Irby, Guerrero, Huerta, and Fan (2012) (Appendix I.1) defined literacy-infused science as reading to learn in science with specific reading/writing skills embedded in instruction and curriculum. Third grade literacy-infused science was tested as a one-year intervention in ELLA-V. **Absolute Priority 1—Moderate Evidence.** VP is based on *moderate evidence of effectiveness* from Project MSSELL (Appendix I.2); Tong et al. (2014), vetted by *WWC as meeting standards without reservations*) ***and from ELLA-V*** for third grade in science from the double-vetted evaluation which meet WWC standards (not yet reviewed), first by the JHU external evaluators and critically reviewed by the ABT Associates with overarching oversight on i3 grants (Appendix I.3). **Absolute Priority 2—Promoting STEM Education.** VP will scale the science education from ELLA, MSSELL, and ELLA-V comparing virtual to face-to-face learning.

A. SIGNIFICANCE **Problems and Potential Contributions.** Few researchers have broached the subject of elementary school students' learning online, particularly with science, and in fact, it was not until April-May, 2020, that the significance of such a study as VP would have been deemed extremely relevant. During this period, teachers reported declining student engagement; spending less time teaching new material, especially in economically challenged schools; and continued problems with access and connectivity (Herold & Yettick Kurtz, 2020). Kurtz (2020) reported in April, 2020, 1,720 educators indicated they were most concerned about virtual learners falling behind in math (90%), English/language arts (88%), and *science* (81%) with similar NWEA study findings related the COVID slide (Kuhfeld & Tarasawa, 2020). *At the time of submission and even after this pandemic, where we may see yet again other viruses emerge (Osaka, 2020), we submit VP to aid in increased evidence-based knowledge under controlled conditions testing the impact of online literacy-infused science education*. Data will be reported after the first year of the study on impacts of online vs. face-to-face teaching and learning. As a rigorous RCT study, VP is significant for increasing knowledge of online delivery effectiveness, since according to the U.S. Department of Education in a meta-analysis by Means, Toyama, Murphy, Bakia, and Jones (2010): "*Few rigorous research studies* of the effectiveness of online learning for K-12 students have been published. A systematic search of the research literature from 1994 through 2006 found *no experimental or controlled quasi-experimental studies* comparing the *learning effects of online versus face-to-face instruction for K-12 students*" (p. xiv). Because the aforementioned study is 10 years old and because the great majority of estimated effect sizes in the meta-analysis are for undergraduate and older students, ***not for elementary*** or secondary learners, there were basically no significant findings related for K-12 students. Chauhan (2017) too noted that few researchers had focused specifically on *elementary students*. It is not only the online elementary learning environment that has not been studied, it is also the *parental/family involvement or engagement with their children in online learning* that is lacking. ***STEM Education Evidence-Based Intervention.*** Related to science which VP is about--only two published papers appeared in a search of the ERIC database on *rural schools and elementary school science and STEM* on the following topics: (a) the use of robotics to generate student interest in math and science (Matson, Deloach, & Pauly, 2004) and (b) teacher, student, and parent attitudes about learning science (Farris-Berg & Project Tomorrow, 2008); therefore, *studies on **science learning in elementary rural school*** are lacking-- VP will add to this body of research. We found ***no experimental and quasi-experimental design studies*** on EC and EL students in rural schools pertaining to low-income schools. ***Professional development (PD) studies** for teachers in rural elementary schools and related to online delivery of instruction are missing* in the literature--VP will rely on Virtual PD (VPD). Although researchers have conducted RCTs exploring teacher PD impact on student science and literacy (e.g., Cervetti et al., 2006; Llosa, et al. 2016; Maerten-Rivera et al., 2016 Palincsar & Magnusson, 2001), we could not locate any such studies in rural schools-- VP will add to the rural science research and practice. ***STEM Interest Analysis*.** In Potvin's and Hasni's (2014) systematic literature review, they indicated that student interest in science declines over the course of K-12 education and that this drop is sharper in *rural and pedagogically traditional* classrooms, with many school factors influencing student interest in science, including enthusiastic, engaging teachers; use of hands-on, inquiry-based learning; and laboratory experiments. No studies were found related to the ***science interest of ELs and ECs in rural elementary schools**, particularly related to **online learning***--VP will add to the literature. ***Virtual Classroom Observation Evidence-Based Scalability.*** While extensive literature has been devoted to systematic observation in English-only classrooms, only a few studies have been conducted exclusively with ELs but none with ELs or ECs in rural elementary classrooms. Garza, Huerta, Spies, Lara-Alecio, Irby, and Tong (2018) emphasized the importance of using a classroom observation instrument to analyze science instruction for ELs, to understand what is working. Lack of a valid, reliable observation instruments can mask the actual classroom practice with teachers who work with ECs and ELs- VP will have an inclusive instrument. Additionally, for virtual classroom observations, *missing from the research and practice landscape is a deep learning (DL) transformer architecture structure with machine learning (ML) related to specific, refined, reliable, and valid pedagogical feedback for teachers' instruction in face-to-face and online observations*. Our research team already has a theoretical and research-based, low-inference observation tool that has been validated across over 1,000 classrooms. A combined team of educational researchers and an external ML industry team will come together in VP to refine an *artificial intelligence (AI) platform that can yield observed real-time teaching practices and provide reliable and valid, timely feedback (online and face-to-face) for scaling nationally in schools and in teacher preparation programs*. ***Innovations Synthesized.*** VP will be implemented *under new conditions* with promising, innovative strategies and proven exceptional approaches with moderate evidence, specifically, an RCT will be implemented (a) across *rural schools* serving high-needs elementary students, including *EC and EL populations*; (b) with *treatment online science instruction and learning compared to control face-to-face* science instruction and learning to determine the degree of impact of virtual innovations on students' *science achievement and reading/writing literacy skills, on parent engagement and interaction, and university science student mentors*; and (c) facilitate scale across a broad rural geographic region by studying classroom observation by human coders and *AI classroom observation technology,* but *it has not yet been attempted* as an effective observation instructional improvement strategy.

B. QUALITY OF PROJECT DESIGN Appendix I.4 includes the longitudinal model and the capacity for teachers and academic support for students and families face-to-face and online are sustainability study potential; additionally, VP has two *levels of implementation*: teacher and student (presented in Appendix I.5). These include: (a) implementation in rural schools across Texas in GRs. 3, 4, and 5, (b) teacher VPD, (c) student-engaged, standards-aligned curriculum that includes the development of academic language in science content via the tested curriculum of Content Reading Instruction Science for English Language Acquisition (CRISELLA, Appendix I.6); (d) a tested virtual mentoring and coaching model (VMC; Appendix I.7) with online delivery, in real time with no delayed feedback, (e) inexpensive technology for teacher VPD (Appendix I.8-I.9), (f) tested virtual observations (VOBS; Appendix I.10) in the classroom and online with a platform for observing and the testing of machine learning for scaling observations (using Dolby GoToRoom), (g) virtual science writing notebooks for Written and Academic oral language Vocabulary development in English in Science (WAVES; Appendix I.11), (h) Family Involvement in Science (FIS; Appendix I.12) with virtual engagement and observation methods, (i) Technology Infusion for English Literacy Advancement in Science (TIELAS) with Nearpod, (j) tested Scientists as Role Models and Mentors (SRM^2), which connects university science majors as mentors to grade-level students, and (i) virtual, online delivery of CRISELLA for in-home instruction. Students will receive 40 minutes of daily science instruction for nine weeks in both delivery mode types, and the teachers will receive 15 weeks of VPD. **CRISELLA** is literacy-infused science curriculum with components to facilitate student reading, comprehension, and development of academic science concepts, following the 5E hands-on science model (Bybee, 1987). Teachers will provide direct instruction (pre-teach pronunciation of 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) in which they process science content as they predict, record, organize, draw, question, and reflect and will benefit from **TIELAS** (equipping classrooms with instructional technology, e.g., teacher and student tablets and Nearpod). **FIS** uses the student tablets to capture interaction between the parent and student during at-home science activities (physical packets will be sent home for C groups, and virtually for online delivery groups; available in both English and Spanish). **SRM^2** is designed to get students motivated about STEM with a virtual mentoring program involving TAMU science majors. **VPD** will be delivered through high-definition, secure video conferencing (GoToMeeting) with a focus on student learning, content, instructional strategies, building capacity for science teaching, previewing upcoming lessons, viewing modeling videos, conducting inquiry activities, and reflection on student learning and teaching practices (T teachers receive PD on building online delivery instructional capacity). Each session will be recorded and links will be sent out to participants so they can have access to review. T and C teachers will also participate in **VMC** using our Applied Pedagogical Education Xtra Imaging System (**APEXIS**) hardware platform with tablets provided by CRDLLA and the GoTo online platform for teacher mentoring and coaching. A trained coach uses APEXIS (we are adding Dolby GoToRoom to this suite of observation tools to refine AI/MS) and virtually observe teacher instruction and offer real-time, bug-in-ear feedback. Teachers complete reflections (use of eduReflection App developed by CRDLLA and ELRC) of their own instruction. Tiered coaching support is provided based on the level of fidelity from observations for improvement. **VICTORY-Virsity** is a Massive Open Online Professional Informal Individual Learning

(MOOPIL; Irby, Sutton-Jones, Lara-Alecio, & Tong, 2015; Appendix I.13) where teachers gain just-in-time PD that ranges from 15 minutes one-time to 10 hours with Aggie STEM. (See VP Logic Model in Appendix I.14). ***B.1. Goals, Objectives, and Outcomes are Clearly Specified and Measurable.*** **GOAL:** Determine the impact of virtual and face-to-face science teaching and learning in rural elementary schools via a longitudinal RCT. VP, a mid-phase project, will serve over the 3 years, 60 RLIS campuses with cumulatively 9,180 teachers (180), high-needs 3rd, 4th, and 5th grade students (4,500), and families (4,500) by building science and literacy instructional capacity, improving students' science and literacy performance and interests, and increasing family involvement in science. The four main ***OBJECTIVES***, ***STRATEGIES***, ***OUTCOMES***, and ***MEASURES*** are presented. ***OBJECTIVE 1.*** To examine the impact of scaling an innovation with moderate evidence, literacy-infused science intervention, to be implemented for a nine-week session (inclusive of six weeks of training for a total of 15 weeks), with 60 randomly assigned rural (RLIS) schools to 30 treatment (online synchronous delivery) vs. 30 control (face-to-face delivery). ***Strategy 1.0.*** Implement VPD for the literacy-infused science intervention for both T and C teachers for six weeks prior to the nine-week teaching session and six weeks prior to that session. ***Outcome.*** *Scaled model validated for VPD for 3rd to 5th-GR. rural teachers for implementation online and face-to-face.* ***Measure 1.0.*** For assessing the impact and efficacy of VPD: (a) observe three times per the nine-week session with the Pedagogical Observation Protocol (POP, formerly named TBOP; Appendix I.15-I.18) with 30 T and 30 C teachers; (b) with discourse analysis rubrics for observing the recorded VPD of T and C teachers; (c) with 10 semi-structured focus groups of T and C teachers pre/post VPD annually; and (d) with a document analysis rubric with T and C teachers' reflections using the eduReflection app).

Strategy 1.1. Use LogMeIn GoToMeeting and APEXIS technology for VPD (train teachers in use of the soft/hardware solutions). ***Outcome.*** *Teachers' ease of use with VPD* ***Measure 1.1.*** Assess the software and hardware ease of use via a Qualtrics VPD-Use Survey of the T and C teachers. ***Strategy 1.2.*** Using GoToMeeting, APEXIS, coordinators observe T and C using Science Teacher Observation Record (STOR; Appendix I.19) for fidelity two times (one initial and mid-way the nine-week period) also used for fidelity of treatment and consequently implement a virtual mentor/coach (VMC bug-in-the-ear program; train T and teachers and coaches on GoToMeeting and APEXIS). ***Outcome.*** *Fidelity of innovation is established, software system has ease of use, VMC model is validated in online instructional delivery.* ***Measure 1.2.*** There are three measures for this strategy: (a) analyze the fidelity of the interventions in the two delivery modes using STOR; (b) assess the software and hardware ease of use via a Qualtrics VMC-Use Survey of the T and C teachers, and (c) with an expert coach, assess, via VMC rubrics, the mentor/coaches' real-time feedback sessions with the T and C teachers and perceived influence of VMC in two modes of instructional delivery. ***Strategy 1.3.*** Develop teachers as reflective practitioners (use eduReflection app). ***Outcome.*** *Teachers who can improve via reflection on their practice.* ***Measure 1.3.*** Data from reflections will be gathered from the eduReflection and analyzed qualitatively for transformation. ***Strategy 1.4.*** Determine the impact of intervention on student achievement and interest between T and C students annually based on their participation in the intervention. ***Outcome.*** *Evidence on student outcomes in science via differing delivery systems (online versus face-to-face) in rural elementary schools.* ***Measure 1.4.*** The State of Texas Assessments of Academic Readiness (STAAR) in science, reading, writing, The Iowa Testing of Basic Skills (ITBS) Science, and Big Idea Science

Assessment (BISA) pre/post, Student Science Interest Survey (SSIS; Appendix I.20), and writing rubrics at 3rd to 5th grades. ***OBJECTIVE 2.*** Determine academic fidelity and sustainability of the science interventions based on the delivery models that occurred with a cohort of 3rd through 5th grade students. ***Strategy 2.0.*** Assess the level of sustainability of the curricular intervention and the instructional delivery model of the cohort of 3rd through 5th grade students in T and C schools. ***Outcome.*** *Scaled model for delivery models in rural schools.* ***Measure 2.0.*** The following measures will collectively assess the sustainability of the curricular intervention and the instructional intervention via the two delivery modes: compare the T and C students' achievement and interests as measured by ITBS Science (GR. 5) and SSIS. ***Strategy 2.1.*** Using the APEXIS technology, coordinators observe T and C teachers and provide feedback on their level of implementation of literacy-infused science and the optimization of sustaining science instructional capacity in online and face-to-face modes, two times per nine week-session with STOR with 30 T and 30 C. ***Outcome.*** *A fidelity instrument, STOR for increasing instructional capacity of science teachers in both online and face-to-face modes.* ***Measure 2.1.*** Analyze the fidelity and sustainability of the innovations using STOR. ***OBJECTIVE 3.*** Provide scaled outcomes with dissemination of products, training, and sustainability beyond the conclusion of VICTORY. ***Strategy 3.0.*** Develop 20 VICTORY-Virsity MOOPILs as an outgrowth for dissemination and scaling, and provide all webinars in a venue called a MOOPIL; VICTORY-Virsity (which will house webinars from T and C teachers, coordinators, mentor/coaches, and principals and online training via Canvas). ***Outcome.*** *MOOPIL site developed and advertised statewide with teachers in the project making presentations sharing information on the curriculum in science and technology, the differences in online and face-to-face teaching and* *tips, and family involvement in science. Micro-credentials awards per number of hours taken.* ***Measure 3.0.*** Pre-post on the MOOPIL modules by 180 teachers who visit the MOOPIL site for ease, use, and relevance for classroom practice. VICTORY-Virsity (virtual targeted PD university) will house multiple webinars for elementary literacy and science teachers across Texas. Assess the number of hits on the VICTORY-Virsity MOOPILs from CRDLLA, ELRC, and Aggie STEM websites. ***Strategy 3.1.*** Determine the differences between the human observations using the low-inference classroom observation tool and the LogMeIn platform for machine learning (AI). ***Outcome.*** *A platform for AI-supported classroom observations that can be scaled for use in elementary schools for in-service teachers and teacher preparation for preservice teachers.* ***Measure. 3.1.*** Data taken from the POP related to language of instruction, communication mode, activity structure, and potentially the level of dense or light cognitive content with comparison of human observations vs. machine learning outputs. ***OBJECTIVE 4.*** To determine the influence of the components of FIS and SRM^2 for families, students, and university science majors. ***Strategy 4.0.*** Using FIS, engage family members and respective T and C students at 3^{rd} through 5^{th} 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. ***Outcome.*** *FIS packages available for school use with suggestions for use in the classroom with take-home packets and with online teaching with the FIS packets being online.* ***Measure 4.0.*** FIS observation rubric (Appendix I.21) already developed and tested to assess academic language use, misconception discussions in science, level of engagement, and student-level vocabulary improvements over time; video analysis input in Atlas.ti; 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 and C students) based on intervention from the university science majors, and (b) T and C students' and university science majors' perspectives of SRM^2. ***Outcome.*** *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 rural students; interviews with rural science students, their teachers, and the university science majors on their perception of positive and negative aspects of SRM^2. ***B.2. Project Design Will Successfully Address Needs of Target Population***. The VP design will address the needs related to provisions for teachers, students, and families in rural areas (technology, paid internet connections for those who do not have sufficient internet connection, materials for use at home). Teachers will receive VPD and VMC, and university science majors will provide mentoring for teachers and students in science for content. Teachers will receive VPD in MOOPILs, curriculum units will be provided, VOBS will be provided with feedback. Additionally, these schools are noted on the RLIS list and include ELs and ECs with the innovations taking into account literacy infusion for both groups in science. ***B.3. Sustained Program of Research and Development, Ongoing Line of Inquiry.*** The coherent program of RCT research reaches back to the original foundational work from the development of the concept of literacy-infused science/science-infused literacy in 2003 with the onset of an IES RCT longitudinal study (grades K-3), then we conducted a validation study of individual components in another RCT initiated in an i3 in 2013 (K-3). Then the same concept of literacy-infused science was tested in a longitudinal RCT in an NSF grant in 2008 (grades 5-6). An i3 validation study is in effect for GRs. 5-8 in an RCT that was initiated 2016— testing teacher sustainability for GR. 5 and student sustainability through GR. 8. A longitudinal study over three grades in rural schools testing face-to-face instruction vs. online instruction has not been done. Additionally, in the latter three, we have improved the virtual components of VPD, VOBS, FOI, VMC, and SRM². We have also been developing and testing with video analysis FIS via go-vision glasses— but not now due to safety issues with the spread of the virus. The ***substantial addition in this line of inquiry follows in four major ways***: **1.** VP adds the inquiry that is s*olely* ***focused on rural schools*** due to the disparities revealed as indicated during the COVID-19 pandemic and teaching and learning in online delivery. **2.** VP ***adds the testing of the impact of literacy-infused science between a classroom face-to-face setting and an online setting in an RCT at the elementary level*** (grades 3-5). To date, *we have not been able to find an RCT at this grade level that tests the efficacy of online teaching and learning*. There has been a very large *uncontrolled national situation* of what teachers have been doing in teaching and learning online. VP provides a solution for mitigating underpreparedness of teachers for unforeseeable circumstances when virtual delivery may confront us again. **3.** *VP* ***adds the testing of virtual classroom observations (face-to-face and online).*** We are testing the similarities and differences between the two with a low-inference, tested observation tool. **4.** VP ***adds the next step in virtual classroom observations to test human virtual observations with a valid tool and platform as compared to AI/ML virtual observations with the same tool and platform. B.4. Increases Efficiency-Time, Staff, Money, Resources-Improved Results & Productivity.*** **Time, Staff, and Money**: Using the virtual components (VPD, VMC, VOBS, Virtual Advisory Board meetings, SRM²) saves time in travel to the many rural locations across Texas and subsequently money related to the time the staff would be gone from the project in driving and overnight stays. Also, the training of the VMCs, virtual observers, and coders for fidelity (graduate doctoral assistants and coordinators) will be more cost effective than going out to train and hire full-time observers to travel and go into classrooms. We are using state assessments or assessments that can be given by teachers in the classroom or online (monitored by them) so that individual testers will not have to travel and spend time individually testing students. **Resources:** All materials will be provided in a private, password-protected virtual learning management system (LMS) for teachers, families, students, and mentors/coaches. Paper will be conserved as writing will be loaded into the online writing journal for students, and other materials for teaching and learning and coaching will be online. Observations will be online in the web-based platform from Inline Resources. We will use GoToMeeting to schedule and observe teachers, and teachers will be provided with licenses to use. For dissemination of the strategies and research webinars, GoToWebinar can hold up to 3,000 participants. With the introduction of AI/ML, an outgrowth of this grant brings entrepreneurial promise and time-saving mechanisms for classroom observations. GotoRoom-GotoMeeting-LogMeIn (GTM) brings the requisite voice and video clarity within the platform and infrastructure required to support data and ML efforts to work with our TAMU education and Aggie STEM team. Having proven teams and data infrastructure out the gate, relieves VP from a capital-intensive, time-consuming ramp-up and greatly reduces the timeline to proof concept implementation. AI usage in observations will reduce time limitations, fidelity, and variability.

C. STRATEGY TO SCALE Coburn (2003) 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 and is addressed with the implementation of virtual teaching and learning in a controlled study. *Sustainability* with VP will provide evidence-based curriculum and delivery mode findings so schools can maintain or sustain the work as the Education Innovation and Research (EIR) funding dissipates. *Spread* is addressed by PD provided to teachers, as well as with close communications with the superintendent and principals. Policy and procedural changes will be analyzed over the scope of the grant, and policy briefs based on evidence will be provided. *Shift of ownership* will be made in phases through the 3-year life cycle starting with initial implementation or deepening of the innovation (Y1), moving to sustaining and spreading (Y2), and spreading, testing the sustainability, and shifting ownership (Y3). In the Victory Summer Institutes, we will teach the teachers in these districts how to conduct VPD with MOOPILs and Virtual Professional Learning Communities related to teaching literacy-infused science and teaching online— strategies that are found successful. When the innovations and effective strategies are then owned by school personnel by the ending of each year, the innovation will be considered to have been institutionalized. If the educational practices remain effective, widespread diffusion and spread is possible. VP will be sustainable beyond the 3-year scope due to the fact that all curriculum, implementation manuals, materials, and MOOPILs via VICTORY-Virsity will be in place with easy access by school personnel nationally or via the CRDLLA. VP meets *unmet demands and offers scale-up opportunities* related to the processes, products, and strategies found and shared related to the literacy-infused science and the instructional delivery mode. **C.1. Specific Strategy(ies) That Address Particular Barrier(s) to Scale.** Though VP appears to be large and ambitious in scope, the PIs can implement such a project. They have experience in managing multiple large, complex projects, working with schools, and producing/disseminating results.

Table 1. *Barriers and Solutions by the PIs Based on Experience with Large-Scale Projects*

Barriers --VP & Experience	**Solution**
Quick turnaround needed to get IRB, positions hired, recruitment of districts/campuses/teachers, obtain districts' memorandum of understanding.	Have IRB approved by ending November so that all schools can be contacted and MOUs finalized by ending January; have personnel posting completed by ending November with hiring completed by January, Have documentation, processes in place by the start of the school year for the intervention to begin.
Geographic spread (60 campus sites across Texas), working out logistics for consent form collection from 60 sites across the state, maintaining confidentiality with online instruction and observations.	Cultivate a vendor relationship with a courier service to obtain consent forms; have an online meeting with teachers and parents on consent and assent; train teachers in obtaining consent; develop onboarding, testing, curriculum, training, observations virtually with teacher orientations, virtual orientation materials, and clearly outlined teacher expectations. (This will be completed with students and families in August each year.)
Selection of science mentors with background checks needed to work with students.	Place flyers on campus; advertise in Jobs for Aggies; advertise with university content specific science organizations (faculty sponsors and student presidents)
Setting up district network/firewall access so we can access cameras to conduct virtual observations; set up agreements with districts for data pullers for state assessments.	Engage local district/campus technology specialist, principal, and superintendent.
Some teachers are unfamiliar with the technologies and cameras (cameras on, mics charged and paired)	Clear communication, training videos; VPD will be developed over the spring semester for summer/fall training of teachers.

In the RCTs we have conducted, there have 60+ presentations, 35 publications, and 12 dissertations. Therefore, the PIs have experience in facilitating dissemination processes, so they can overcome such barriers. The PIs have recently developed EduReflection, an open access Droid and Apple app that overcomes the and barriers of access for reflecting and teacher observation instruments that are web-based for use. The PIs also have no barriers working with external evaluation teams and Advisory Boards. They have managed large research teams. A partnership with LogMeIn will facilitate the barrier to scaling the AI/ML component of the virtual observations***. C.2. Mechanisms to Disseminate Information for Development or Replication*** The VP components will be replicable for a variety of school types across Texas or the nation at a very reasonable cost. The total cost of such a program is minimal over the three years—approximately $871.41 per person for implementation of VP. Much of the components on "how to" implement such a program on the campus will be available online via MOOPILs. Included in the distribution for replication across rural schools initially (however, evidence-based strategies for online delivery may also apply to non-rural elementary teachers) are the products mentioned in Section B. Traditional papers will be published, along with presentations offered. However, webinars from CRDLLA will be provided and from the grant during the three years. Research Briefs will be published on the Center websites. The AI proof of concept for classroom observation, science-infused literacy, MOOPILs, FIS, virtual components, and the evidence-based strategies for online teaching and learning will be shared.

D. ADEQUACY OF RESOURCES AND QUALITY OF THE MANAGEMENT PLAN

D.1. Capacity to Bring Project to Scale. VP PIs have strong experience and are highly qualified to bring the project to scale. The PIs (Lara, Irby, & Tong) have taken PD to scale with online webinars with over 400 registered for each one; for face-to-face over 500 were registered. They have taken to scale reflection with the eduReflection app; they have taken to scale programs that compact curriculum for in-service teachers and train large numbers online for microcredentials for personalized learning; they have taken to scale evidence-based curriculum and have via the TAMU System commercialized one curriculum from one of the early projects (STELLA) with Frog Street Press (to be marketed and published 2020). There have been many substantial publications (five reviewed by WWC, with two with/without reservations) and presentations.

Continuous improvement has come in the form of communication with district superintendents, curriculum directors, principals, teachers, the evaluation team, and the local project team. The personnel for the grant are: ***Dr. Rafael Lara-Alecio*** (PI, Project Director), has led a team of researchers for over two decades providing the status for the moderate evidence that undergirds VP. He has directed over $80,000,000 in grants. ***Dr. Beverly Irby*** (Co-PI) has also led research teams and has been responsible for curriculum development and family engagement components and micro-credentials with VPD, VMC models, and the Reflection Cycle. She has managed $18,000,000 in grants over the past three years. ***Dr. Fuhui Tong*** (Co-PI) has led research teams and trained graduate students in evaluative data collection online, with teleform, and with statistical procedures. She has been a Co-PI on large RCT grants and teaches evaluation at the doctoral level and has led grants as well. ***Dr. Robert Capraro*** (Co-PI) has led many grant and contract programs for Aggie STEM which he directs. His research team works closely and in the same space as our Centers' research teams. He is a prolific author and trains graduate students in research and development. ***Dr. Mary Margaret Capraro*** (Co-PI) is Co-Director of Aggie STEM. She is an expert in STEM curriculum and making connections across the four areas and with literacy. She is in charge of Aggie STEM Camp each summer at TAMU for teachers and for students and works with districts and schools across Texas. ***Major CRRE Evaluation Personnel. Dr. Steven Ross*** will lead the randomized control trial of the program. He is currently senior research scientist and professor at the CRRE, 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. ***Dr. Rebecca Wolf*** has expertise in educational program evaluation, quantitative research design (including RCT, quasi-experiments, and mixed methods), statistical methods, and data management. **<u>Resources for Scaling VP</u>.** Within the grant budget, there is sufficient person-power to manage the grant. There is a lead coordinator, post-doc first year, two other content coordinators, one logistics coordinator, six graduate assistants and two student workers. There are two healthy Centers, CRDLLA and ELRC, that have undergone recent positive external evaluations; they share an Outreach Coordinator who will be working with us as well. We have a new building provided by the University that is adequate for all personnel and is well-equipped. We have worked with and have access to a full service support staff with attorneys for entrepreneurship and commercialization at the TAMU System. The evaluation will be conducted at JHU, one of the nation's premier research institutions. CRRE will provide strong and consistent organizational support. ***<u>Support and Commitment.</u>*** 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. ***<u>Resources.</u>*** JHU's, and our, facility provides office admin services; photocopying and conference rooms; and areas for maintaining/analyzing data ***<u>D.2. Costs are Reasonable in Relation to Objectives, Design, and Significance.</u>*** At the end of VP, we will have served a total of 9,216 participants across three years. Of the total federal funds, 16.7% are allocated for direct participant support costs to serve participants in direct support. The remaining funds are budgeted for project participant support and implementation costs, including the participants' PD costs, technology supplies, curriculum supplies, and personnel support, data collection/analysis, and evaluation of VP. The entire project has the potential to be replicated for similar districts, schools, teachers, parents and families across the country. The proximal cost for participants served over the three-year project period is approximately $871.41. ($7,999,563/9, 180 participants—teachers, students and parents/families). MOOPILs, via VICTORY-Virsity, have the potential to eventually reach all elementary teachers across RLIS communities. Costs are reasonable in relation to objectives, design, and significance for enhancing quality face-to-face and online science education in RLIS districts in Texas. ***D.3. Potential for Continued Support of Project After Funding Ends*****.** Sustainability, replication, and benefits by the different components were addressed before. The long-term sustainability of our program is a high priority for the Texas A&M Research Foundation, CRDLLA, and ELRC. We have plans to work with the TAMU Technology Commercialization and the private industry platform with LogMeIn to sustain an AI solution in the educational arena for observations for teachers in two delivery modes. We have plans as well to commercialize some of the MOOPILs based on the concept that we build for this program. The curriculum for this project overall will be made available; however, the training for schools will be provided via the Center for a fee for service. Our funding strategies include building relationship with other foundations, cultivating support from corporate sponsors, increasing revenues from high quality institutes and webinars, and refining our websites and keeping them current. The sustainability and incorporation of the activities and benefits are proposed in anticipation of positive results from rigorous research. ***D.4. Adequacy of Management Plan to Achieve Objectives on Time Within Budget.*** There are four management groups involved in VP. 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 Texas A&M Research Foundation and the CEHD Post-Award Research Office. The PIs will train coordinators, work with the external evaluation team in all evaluation components, and will hire the implementation personnel. The *Application/ Implementation Group (AIG)* is made up of coordinators, research associates, 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 is charged with gathering data and implementing the evaluation design and analysis described in this application. Table 2 includes the major management milestones with objectives and strategies, group responsible, and timeframe. The milestone chart will be kept on Google and will be discussed monthly with the team.

Table 2. *Major Milestones, Group Responsible, and Timeframe*

Major Milestones	**Responsible Group**	**Y 1**	**Y 2**	**Y3**
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**	√	√	√
Objective 2.0				
Strategy 2.0	**EG**			√
Strategy 2.1	**EG; AIG**	√	√	√
Objective 3.0				
Strategy 3.0	**EG; AIG**	√	√	√
Strategy 3.1	**AIG**	√	√	√
Objective 4.0				
Strategy 4.0	**EG; PPO; AIG**	√	√	√
Strategy 4.1	**AIG; EG**	√	√	√

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 Project Officer on Mgt. and GPRAs	PPO	√		
Meet with Project Office monthly	PPO;AIG	√	√	√
Meet with EG four times annually or as needed	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

E. QUALITY OF THE PROJECT EVALUATION ***E.1. Project Will Produce Evidence of Effectiveness to Meet WWC Standards.*** The team has extensive experience designing and conducting evaluations to meet WWC Evidence-Based Standard. The overall VP 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. **Sample and Research Design.** The evaluation includes a clustered randomized design with 60 Texas rural schools randomly selected from the 112 schools in 77 districts that have expressed interest in participation, and randomly assigned (by the EG) to either a treatment (T=30) or control (C=30) condition. The integrity of such assignment will be maintained because when a school is assigned to receive T in GR. 3 in Y1, then this school will continue to receive T in the subsequent years; the same is true for C schools. The majority of these rural schools only have one teacher/class in GRs. 3-5, which is typical for rural schools. For schools that have more than one teacher, one teacher will be randomly selected by the EG for the randomized campus to implement the corresponding condition. The intervention will be provided to a longitudinal student sample, or more specifically, to GR. 3 students and their teachers in Y1, to those same students in GR. 4 and their teachers in Y2, and to those same students in GR. 5 and their teachers in Y3. We used the federal guidelines for the RLIS program to identify eligible districts in Texas and contacted these districts about project participation. EC students in these 77 RLIS districts that have expressed support with letters (Appendix C.3) make up 44% to 100% of the population, and EL students are 0% to 74%. To test the **sustained impact** of this project, as these students complete GR. 5, we will examine the accumulative treatment effect on the domains that are of interest (see Appendix I.22), and as rural 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. **Power Analysis**. In VP, because students will be nested in classrooms, which in turn are nested within schools, we used PowerUp! (Dong & Maynard, 2013) to determine the number of clusters (which are schools in this case). 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 .20 (although the effect size of GR. 3 CRISELLA is .27 in science, and ranges between .35 to .7 in GR. 5 intervention in science and reading measures, we decided to use a more conservative effect size in this study), a cluster size of 25 (including students in one class per school), and ICC of .10 (given that all participating schools are located in Texas rural areas that share similar characteristics) (Hedges & Hedberg, 2007). Using these parameters and taking into consideration potential attrition rate over time, by the end of the project year, we can detect an effect size of 0.20 with a power of over 0.80 if we start with 56 schools. We over-sample to 60 schools to account for potential attrition at school level, and will be able to maintain an effect size of .20 with a cluster size of 20 by the end of the project. **Attrition and Missing Data**. In our previous longitudinal RCTs that included urban and suburban schools, the attrition at the school level was low. In one RCT we were 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 anticipate a low cluster level attrition in VP. At the student level, according to Texas Education Agency (TEA, 2018), student mobility rate in rural schools was about 16%. Therefore, we do not anticipate a high overall attrition rate in VP. Further, the EG will follow the model proposed by What Works Clearinghouse (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. The analysis sample is defined as all cases with non-missing outcome data. **Fidelity of implementation** will be taken on key program components including % of teachers who miss more than two VPD trainings; % of teachers who attend at least one coaching session, and % of schools that receive curriculum materials. The threshold to determine high level of fidelity is determined by the EG to be 90%. **Questions to Evaluate Objectives and Data Analysis per Question. (S=Strategy)** **Confirmatory questions:** **1A.(S1.0).** What is the impact of VP on teachers' instructional delivery as measured by POP? *Analysis: Regression analysis, controlling for round 1 scores will be conducted, with T as predictor variable, and round 3 scores as outcome variable to examine the impact of T on teachers' instructional delivery by each grade level*. **1B.(S 2.0)** What is the impact of VP on students' achievement in science, reading, and writing, and science interest longitudinally from GRs. 3-5, compared with C? *Analysis: we will use hierarchical linear modelling (HLM) to analyze the longitudinal T effects at the end of GR. 5. Student will be the level-1 unit of analysis, with pre-test score as covariate (e.g., ITBS and SSIS); school will be the level-2 unit of analysis. A simple presentation of the model follows:* $Y_{ij} = \gamma_{00} + \gamma_{10} Pretest_{ij} + \gamma_{01} Treatment_j + u_{0j} + e_{ij}$ *. Outcome measures for this confirmatory question will be GR. 5 state assessment in science, reading, and writing, science interest, and science writing notebooks.* **Mediation question**: **1C.(S 1.4)** Is the T effect on students' achievement in science, reading, and writing, and science interest mediated by their teachers' instructional practices annually in GRs. 3-5? *Analysis: We will use HLM to conduct this. Our previous research has identified that T effect was completely mediated through teachers' time allocation in academic English on ELs' English retell fluency (Tong et al., 2017). In that study, it was science-infused literacy intervention. Therefore, we are interested in exploring such mediation effect in VP with literacy-infused science intervention. Because in this study there will be one teacher selected per school, we will examine the upper-level mediation effect in RCT (Pituch & Stapleton, 2012; VanderWeele, 2010), which is teachers' pedagogical practice as measured by STOR and POP. Similar model will be established adding level-2 predictors such as teachers' literacy-infused science instruction to test the mediation effect:* $Y_{ij} = \gamma_{00} + \gamma_{10} Pretest_{ij} + \gamma_{01} Treatment_j + \gamma_{02} TeacherPedagogy_j + u_{0j} + e_{ij}$.

Exploratory and implementation questions: OBJECTIVE 1—**1D.(S 1.0, 1.1, 1.3)** 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 (via eduReflection app). Data analyzed using constant comparative method (Creswell, 2014) and coded according to themes for identifying trends or patterns with all data entered into Atlas.ti software. Focus groups with T teachers at GRs. 3-5 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. **1E.(S 1.2)** To what extent do teachers implement the innovations with fidelity? *Analysis: Analysis of covariance (ANCOVA) will be conducted to compare T and C teachers on STOR, using initial data taken as baseline.* **1F.(S 1.2)** How do teachers perceive the ease of use and quality of VMC using the LogMeIn GoToMeeting and APEXIS software and hardware? *Analysis: (See 1D). Data from VMC rubrics will be analyzed to compare between T and C.* **1G.(S 1.3)** What is teachers' perceived effectiveness of the following components: VPD, interventions, observation tools, and student achievement? *Analysis: (See 1D).* **1H.(S 1.4)** What is the impact of VP on students' achievement in science, reading, and writing, and science interest by grade, compared with C? *Analysis: We will use HLM to analyze the T effects annually in GRs. 3-5. Student will be the level-1 unit of analysis, with pre-test score as covariate (e.g., ITBS and SSIS); school will be the level-2 unit of analysis. A simple presentation of the model follows:* $Y_{ij} = \gamma_{00} + \gamma_{10}Gr3pretest_{ij} + \gamma_{01}Treatment_j + u_{0k} + e_{ij}$. **1I.(S 1.4)** Do student characteristics (e.g., language status, socio-economic status, and ethnicity) predict their science, reading, writing achievement, and science interest annually and longitudinally from GRs. 3-5?

Analysis: An exploratory subgroup analysis will be conducted to include student characteristics (EL/EC/ethnicity) as student-level predictors to explore the impact of these variables on their achievement. We will use the same models specified in 1G and 1H, adding student level predictors in these models. **OBJECTIVE 2** **2A.(S 2.1)** Do teachers' instructional practice improve as a result of VPD as measured by STOR? *Analysis: Paired-sample t-test will be conducted to identify the improvement on STOR by grade.* **OBJECTIVE 3**— **3A.(S 3.0)** How do teachers who access VICTORY-Virsity benefit from taking MOOPILs in relation to literacy-infused science curriculum, technology, tips on virtual teaching, and family involvement in science? *Analysis: Paired-sample t-test will be conducted to identify the change before and after teachers taking each of the 20 MOOPIL sessions.* **3B.(S 3.1)** Is there a difference on teachers' instructional practices as coded by human observation and by the AI? *Analysis: We will calculate agreement coefficient of AC_1 (Gwet, 2014) in each category of the observation protocol to examine alignment between AI and human gold standard coding.* **OBJECTIVE 4**— **4A.(S 4.0).** In what ways do family members in GRs. 3-5 engage in science dialogue through FIS related to (a) use of academic language, (b) use of misconceptions in science, and (c) level of engagement from the parent's and child's perspectives? *Analysis: chi-square analysis will be conducted to test the difference between T and C families on time allocation in the above areas using observation rubrics developed from an earlier project for video analysis of family recordings. This question will also be analyzed qualitatively (see 1D).* **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' attitudes toward science increase based on the FIS activities? *Analysis: (See 1D, with survey developed in a previous project on attitudes* *and satisfaction of students and parents toward science). Paired-sample t-test will be conducted to identify the improvement on attitudes and satisfaction before and after FIS.* **4C.(S 4.1)** How do university science majors engage with rural school students? *Analysis: Descriptive statistics will be conducted, and qualitative analysis based on notes taken from VOBS will also be performed (See 1D).* **4D.(S 4.1)** What are university science majors' and the rural school teachers and students' perceptions on SRM^2? *Analysis: (See 1D).* ***E.2. Plan Clearly Articulates Components, Mediators, Outcomes, and Measurable Threshold of Outcomes, & E.3. Validity and Reliability of Outcome Measures.*** We will compare students' achievement on the constructs aligned with the research questions as measured by three assessment types (Appendix I.22). First, we will administer researcher-developed writing rubric (Huerta et al., 2014, Appendix I.23) to measure and compare the academic language development and conceptual understanding through science notebooks. The reliability of the writing rubric is 0.89 for language and 0.88 for science (Hurta et al., 2014). The second is the Big Ideas in Science Assessment (BISA, developed in MSSELL with an internal consistency of .79, Lara-Alecio et al., 2018). BISA is curriculum-based and formative assessments aligned with state and national standards and are embedded in instruction to provide timely feedback in order to adjust instruction to improve learning. The third assessment is Student Science Interest Survey (SSIS, with a reliability of .86, Tong et al., 2019). The second group of assessments is the state-mandated, standards-aligned STAAR assessments, including science (GR. 5), reading (GR. 3-5), and writing (GR. 4). STAAR measures academic progress of all students, including ELs and ECs. According to Technical Digest 2018-2019, the reliability of STAAR GRs. 3-5 Reading ranges from 0.89 to 0.90, GR. 4 writing 0.85, and GR. 5 science 0.88. (TEA, 2019). Finally, we also plan to administer a rigorous standardized science instrument to measure progress in science, the Iowa Test of Basic Skills (ITBS)-Science to GRs. 3-5 students. ITBS science assesses not only students' knowledge of scientific principles and information, but also the methods and processes of scientific inquiry. The reliability of ITBS-Science Form E levels 9-12, for grades 3 to 6, ranges from 0.84 – 0.87 (Kuder-Richadson Formula 20 [K-R 20], Dunbar et al., 2015). 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., BISA and science notebooks); and *distal level* assessment based on state or national standards in a particular domain (i.e., STAAR and ITBS). In addition, teacher-level data will also be collected from POP and STOR for confirmatory and mediation analyses. These two instruments were both used in the second study of moderate evidence and are considered as fidelity of intervention. STOR is an instrument 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. VOBS using POP will be conducted three times across the nine-week session in both T and C classrooms also as part of the fidelity measure. POP will be adapted from TBOP (Lara-Alecio, Tong, Irby, & Mathes, 2009) to capture literacy-infused science instruction. A rigorous process has been established and reported in the training and monitoring of inter-rater reliability of STOR and TBOP (Tong et al., 2019) and will be applied to POP. Inter-rater reliability using Gwet's (2012) AC_1 coefficient is reported to range from .724 to .945 (Tong et al., 2020). **Data collection** by the EG will occur in Y1-3. See Appendix I.22 for a summary of measures for planned contrasts and data collection.

References

Burke-Johnson, R. (1997) Examining the validity structure of qualitative research. *Education, 118*(2), 282-292.

Bybee, R. W. (1987). Science education and science-technology-society (S-T-S) theme. *Science Education, 71*(5), 667-683.

Cervetti, G., Pearson, P. D., Bravo, M. A., & Barber, J. (2006). Reading and writing in the service of inquiry-based science. In R. Douglas, M. Klentschy, & K. Worth (Eds.), *Linking science and literacy in the K-8 classroom* (pp. 221-244). Arlington, VA: NSTA Press.

Chauhan, S. (2017, February). A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education, 105*: 14-30.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3-12.

Creswell, J. W. (2014). Research Design Qualitative, Quantitative, and Mixed Methods Approaches (4th ed.). Thousand Oaks. CA: SAGE Publications.

Dong, N., & Maynard, R. A. (2013). *PowerUp!*: A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs. *Journal of Research on Educational Effectiveness*, *6*(1), 24-67. doi: 10.1080/19345747.2012.673143

Dunbar, S., Welch, C., Hoover, H. D., Forsyth, R. A., Frisbie, D. A., & Ansley, T. N. (2015). *Forms E and F: Research and development guide*. The University of Iowa. Retrieved from https://itp.education.uiowa.edu/ia/documents/Research-Guide-Form-E-F.pdf

Farris-Berg, K., & Project Tomorrow. (2008). *Inspiring the next generation of innovators: Students, parents and educators speak up about science education*. Retrieved from https://tomorrow.org/speakup/scienceReport.html

Freidhoff, J. R. (2017). *Michigan's K-12 virtual learning effectiveness report: 2015-2016*. Lansing, MI: Michigan Virtual University. Retrieved from https://media.mivu.org/institute/pdf/er_2016.pdf

Garza, T., Huerta, M., Lara-Alecio, R., Irby, B. J. & Tong, F. (2018). Pedagogical differences during a science and literacy integrated intervention for English language learners. The Journal of Educational Research, 111(4), 487-496. doi: 10.1080/00220671.2017.1302913

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (4th ed.). Gaithersburg, MD: Advanced Analytics.

Hedges, V. L., & Hedberg, C. E. (2007). Intraclass correlations for planning group randomized experiments in rural education. *Journal of Research in Rural Education, 22*(10), 1-15.

Herold, B., & Yettick Kurtz, H. (2020, May 11). Teachers work an hour less per day during COVID-19: 8 key EdWeek survey findings. *Education Week*. Retrieved from https://www.edweek.org/ew/articles/2020/05/11/teachers-work-an-hour-less-per-day.html?intc=main-mpsmvs

Irby, B. J., Sutton-Jones, K. L., Lara-Alecio, R., & Tong, F. (2015). Informal individual learning via virtual professional development: A proposal for massive open online professional informal individual learning (MOOPIL). In F. Muyia Nafukho & B. J.

Irby (Eds.), *Handbook of research on innovative technology integration in higher education* (pp. 343-355). Hershey, PA: IGI Global.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.

Kuhfeld, M., & Taraswa, B. (2020, April). *The COVID-19 slide: What summer learning loss can tell us about the potential impact of school closures on student academic achievement*. NWEA. Retrieved from https://www.nwea.org/content/uploads/2020/05/Collaborative-Brief_Covid19-Slide-APR20.pdf

Kurtz, H. (2020, May 10). National survey tracks impact of coronavirus on schools: 10 key findings. *Education Week*. Retrieved from https://www.edweek.org/ew/articles/2020/04/10/national-survey-tracks-impact-of-coronavirus-on.html

Lai, H. C., Tong, F., Yoon, M., Lara-Alecio. R., Irby, B. J., & Kwok, O. (2014, April). *Modeling cross-classified data with adequate within-subject variance-covariance structures: A lesson from the Project ELLA data*. Paper presented at the annual meeting of American Educational Research Association, Philadelphia, PA.

Lara-Alecio, R., & Parker, R. (1994). A pedagogical model for transitional English bilingual classrooms. *Bilingual Research Journal, 18*(3&4), 119-133.

Lara-Alecio, R., Tong, F., Irby, B.J., Guerrero, C., Huerta, M, & Fan, Y. (2012). An experimental study of science intervention among middle school English learners: Findings from first year implementation. *Journal of Research in Science Teaching*, *49*(8), 987-1011. doi: 10.1002/tea.21031

Lara-Alecio, R., Tong, F., Irby, B. J., & Mathes, G. P. (2009). Teachers' pedagogical differences among bilingual and structured English immersion kindergarten classrooms in a randomized trial study. *Bilingual Research Journal, 32*(1), 77-100.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.

Llosa, L., Lee, O., Jiang, F., Hass, A., O'Connor, C., Van Booven, C. D., & Kieffer, M. J. (2016). Impact of a large-scale science intervention focused on English language learners. *American Educational Research Journal, 53*(2), 395-424.

Maerten-Rivera, J., Ahn, S., Lanier, K., Diaz. J., & Lee, O. (2016). Effect of a multiyear intervention on science achievement of all students including English language learners. *The Elementary School Journal, 116*(4), 600-624.

Matson, E.; DeLoach, S., & Pauly, R. (2004). Building interest in math and science for rural and underserved elementary school children using robots. *Journal of STEM Education: Innovations & Research,* 5(3/4), 35-46.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. Retrieved from https://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf

Mehta, J. (2013a). The allure of order: High hopes, dashed expectations, and the troubled quest to remake schooling. New York: Oxford University Press.

Mehta, P. (2013b). *xxM user's guide*. Retrieved from http://xxm.times.uh.edu/downloads/xxm.pdf

Osaka, S. (2020). These scientists saw COVID-19 coming. Now they're trying to stop the next pandemic before it starts. *Mother Jones*. https://www.motherjones.com/environment/2020/05/these-scientists-saw-covid-19-coming-now-theyre-trying-to-stop-the-next-pandemic-before-it-starts/

Palincsar, A. S., & Magnusson, J. (2001). In S. Carver & D. Klahr (Eds.), Cognition and instruction: Twenty-five years of progress (pp. 151-194). Mahwah, NJ: Erlbaum.

Peele, H., & Riser-Kositsky, M. (2020, March 6). Map: Coronavirus and school closures (2020, March 6). *Education Week*. Retrieved from https://www.edweek.org/ew/section/multimedia/map-coronavirus-and-school-closures.html

Pituch, K. A, & Stapleton, L. M. (2012). Distinguishing between cross- and cluster-level mediation processes in the cluster randomized trial. Sociological Methods & Research, 41, 630–670. doi:10.1177/0049124112460380

Potvin, P. & Hasni, A. (2014). Interest, motivation, and attitude towards science and technology at K-12 levels: a systematic review of 12 years of educational research. *Studies in Science Education, 50*(1), 85-129.

Raudenbush, S. W., Liu, X. F., Spybrook, J., Martinez,A., & Congdon, R. (2006). Optimal Design software for multi-level and longitudinal research (Version 1.77) [Computer software]. Retrieved from http://sitemaker.umich.edu/group-based

Ruiz-Primo M. A., Shavelson, R. J., Hamilton, L. S., & Klein, S. (2002). On the evaluation of systematic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching, 39*(5), 369-393. doi:10.1002/tea100027.

Song, M., & Herman, R. (2010). Critical issues and common pitfalls in designing and conducting impact studies in education: Lessons learned from the What Works Clearinghouse (Phase I). *Educational Evaluation and Policy Analysis, 32*(3), 351–371.

Texas A&M University. (2016). High impact practices [Webpage]. Retrieved from https://sllo.tamu.edu/highimpactpractices/

Texas Education Agency (2018). *2017–18 Texas academic performance reports*. Retrieved from https://rptsvr1.tea.texas.gov/perfreport/tapr/2018/index.html

Texas Education Agency (2018). *Campus and district type data search (2017-2018*). Retrieved from https://tea.texas.gov/reports-and-data/school-data/campus-and-district-type-data-search

Texas Education Agency. (2019). *Enrollment in Texas public schools, 2018-19.* (Document No. GE19 601 13). Austin, TX: Author. Retrieved from https://tea.texas.gov/sites/default/files/enroll_2018-19.pdf

Texas Education Agency. (2019). 2019 STAAR mean p-values and consistency values by reporting category and content area. Austin, TX: Author. Retrieved from https://tea.texas.gov/sites/default/files/digest19-appendB-STAAR-4-Reliability.pdf

Tong, F., Irby, B. J., Lara-Alecio, R., Guerrero, C., Fan, Y., & Huerta, M. (2014). A randomized study of literacy integrated science intervention for low SES middle school students: Findings from first year implementation. *International Journal of Science Education*, *36*(12), 2083-2109. doi: 10.1080/09500693.2014.883107.

Tong, F., Luo, W., Irby, B. J., & Lara-Alecio, R., & Rivera, H. (2017). Investigating the impact of professional development on teachers' instructional time allocation and English learners' cognitive and academic language development: A multilevel cross-classified approach. *International Journal of Bilingual Education and Bilingualism*, *20*(3), 292-313. doi: 10.1080/13670050.2015.1051509

Tong, F., Tang, S., Irby, B. J., Lara-Alecio, R., & Guerrero, C. (2020). Determining proper indices of inter-rater reliability of classroom observation instruments as fidelity measure in large-scale randomized research with English learners. *International Journal of Educational Research, 99*. https://doi.org/10.1016/j.ijer.2019.101514

Tong, F., Tang, S., Irby, B. J., Lara-Alecio, R., Guerrero, C., & Lopez, T. (2019). A process for establishing and maintaining inter-rater reliability for two observation instruments as fidelity of implementation in a large-scale randomized controlled trial. *Studies in Educational Evaluation*, *62*, 18-29. doi: 10.1016/j.stueduc.2019.04.008

VanderWeele, T. (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological Methods & Research, 38*, 515–544. doi:10.1177/0049124110366236