Table of Contents

This mid-phase study is designed to evaluate whether culturally responsive project-based science learning proactively impacts high-need rural secondary school students in the southern region of the US. Absolute Priorities 1 and 3 and the competitive preference are outlined below.

Absolute Priority 1 – Moderate Evidence. This project implements and evaluates **Crafting Engaging Science Environments** (CESE), an innovative science intervention for chemistry and physics high school students shown to have moderate evidence, to a new population of high-need students in US southern rural schools. CESE has been piloted, field-tested, and rigorously tested in Michigan and California with an efficacy RCT of several thousand diverse students ([1]). This study showed that students in the treatment condition significantly outperformed the control students on a state standardized science test (0.20 effect size, which is considered a large treatment effect; [2]) and increased college ambitions. Study results provide evidence that CESE would meet the requirements of the *What Works Clearinghouse (WWC) Standards* with reservations (see EIR Proposal Evidence Form), which establishes the intervention as having moderate evidence meeting absolute priority 1.

Absolute Priority 3 – Promoting Equity in Student Access to Educational Resources and Opportunities: STEM. The CESE intervention was designed to assist all students in succeeding in science courses that often serve as gatekeepers, posing barriers to college admission especially for high-need students [3]. Focusing on what students should know and be able to do, CESE was based upon the principles of project-based learning [4,5,6,7] three dimensions of learning articulated in the *Framework for K-12 Science Education* [8], and performance expectations recommended in the Next Generation Science Standards [9]. One of the few new project-based secondary science programs aligned with the NGSS [10,11], CESE serves all high school students, not just those in advanced placement or honors programs. Created to advance a new vision of science teaching and learning, CESE consists of evidence-based teacher and student materials including technologically incorporated activities [12,13,14,15], end-of-unit formative assessments, and professional learning supports. CESE also enhances engagement with meaningful and challenging science activities [16].

This proposal is designed to learn if CESE can be replicated with a new population of high-need students, who attend public schools in the rural South, where one in six live below the poverty line, one in seven qualify for special education, and one in nine have moved at least once in the past 12 months. Alabama and North Carolina, where this intervention will be located, were tied for second among the top highest-priority states faced with multiple education challenges in a recent ranking [17].

Competitive Preference: Implementers and Partners. Historically Black Colleges and Universities (HBCUs) are recognized as building successful gateways to higher education and the workforce, especially for students of color, low-income and first-generation college goers [18] Given the importance and expertise of HBCUs in advancing science education, Michigan State University (MSU) formed a new partnership with Alabama A&M University (AAMU) and Winston Salem State University (WSSU) to implement the CESE intervention to a new population and to ensure that revisions to the curricula are culturally responsive for students in public secondary schools and particularly those in local census codes (32, 33, 41, 42, and 43). Results from this collaboration will capitalize on the unique strengths of our partners’ knowledge and experience as we work together revising and implementing the CESE curriculum, helping students connect science and engineering concepts and principles to questions that are meaningful to their lives.

The aim of the MSU, AAMU, and WSSU partnership is to provide culturally responsive high quality instructional materials and professional learning to science teachers in diverse school locations in the south (i.e., with an oversample of schools over 51% black and in rural areas); conduct a field test with new modifications to the CESE intervention; and test its effectiveness with an independent evaluator employing a randomized control trial (RCT).

I. Significance

I.1 Current Issues of science education in the South. Unquestionably, the importance of science and engineering are of deep concern to our society's health, economic development, emerging technologies, and civic participation [19]. However, US students are not keeping pace with the knowledge, skills, and technology identified as critical for the future. The latest *National Assessment of Education Progress Report Card* [20] showed that the scores of 8th and 12th graders were not significantly different than in 2015 and the majority of 12th grade students are at or below proficiency. Predicted post-Covid-19 results are estimated to be to even lower especially for high-need students [21,22]. These low levels of science proficiency are indeed problematic as they highlight a ten-year trend in science performance that must stop. Science learning is an essential education goal and must be supported with rigorous interventions [22].

The situation for students in the rural south is particularly problematic. The recent report on our nations' rural schools [23] shows that in Alabama although nine out of 10 students from rural districts graduate from high school, fewer than five percent of students earned college credit compared to rural students in other states. In North Carolina, rural students are at or below the national median on estimated graduation rates, dual enrollment, and taking college admission tests (all these estimates were produced prior to the pandemic).

Implementing CESE will improve students' science academic performance, keep students on track to successfully complete physical science gatekeeper courses, and strengthen their academic portfolio for admission to a two- or four-year higher education institution or technical school. With its engaging and challenging curriculum, the intervention is purposively designed to also encourage, support, and provide social, and emotional science learning opportunities that have short-term outcomes such as a willingness to take on challenges [24] and increased confidence in solving real world scientific problems with enhanced knowledge and skills [25] and long-term interest in continuing science education after high school graduation [26].

I.2 Rationale and Significance of CESE for Academic, Social and Emotional Learning. Over the last twenty years, a number of science reports [27, 28, 29, 30] have been undertaken that raise the importance of equitable access to science and engineering teaching and learning for today's students. The CESE intervention is grounded in two of the most important of these national reports that explain how science learning needs to be reformed. Briefly, the first of these reports was, *A Framework for K-12 Science Education*, (*Framework*) by the National Research Council created in 2012, describing a vision of science learning that focuses on solving real world problems through three-dimensions of scientific knowledge, which include science and engineering practices (SEPs), crosscutting concepts (CCCs), and disciplinary core ideas (DCIs). SEPs are behaviors scientists perform as they build theories about natural phenomena through investigations, creating models, generating explanations, and science-based arguments. CCCs refers to ideas linked and found across disciplines, which integrate knowledge from multi-disciplines to explore phenomena. DCIs center on the major ideas of a science discipline. These three-dimensions of scientific knowledge are viewed as critical for investigations and problem-solving opportunities, connecting to scientific concerns and technical knowledge, and expanding in depth and complexity across multiple grade levels [31]. Closely following the release of the *Framework*, was the Next Generation Science Standards [32], which provides a set of standards for science learning that describe performance expectations which identify what students should know and be able to do. These two reports describe how science learning should be taught, what students should know, and the experiential activities that support learning opportunities and the rationale upon which the innovative strategy for CESE has been developed.

Despite the widespread adoption of NGSS-like standards, there is a lack of research on evidence-based curricula and lessons that align with NGSS and the NRC recommendations. CESE was initiated to fill this gap by designing a system approach that exemplified the *Framework*, incorporated NGSS performance standards for high school chemistry and physics, and created instructional opportunities where students used their lived experiences to ask meaningful questions investigate phenomena, design solutions, and create artifacts in collaboration with their classmates. To meet these goals, the team turned to the project-based learning (PBL) principles designed by [redacted] [33]. Incorporating these principles, CESE was created not as a stand-alone curriculum but as a system approach which includes learning activities and materials for students and teachers, assessments, and professional learning experiences [34].

Many students, especially those in the rural south, have not been exposed to many scientific advances in curriculum and technology [35, 36]. What makes CESE so valuable to this population is that its most fundamental principle for engaging students is having them ask meaningful questions from their own lives as they begin investigating and explaining the causes of phenomena and designing evidence-based solutions [37]. By having students ask questions that reflect on their own lives and make meaningful connections between themselves and the scientific world, the current version of the curriculum allows them to build on their own experiences; however, the curriculum requires revision to be responsive to their unique cultural experiences and ideas.

1.3 Creating a CESE Culturally Responsive Curriculum. The culturally responsive framework we have adapted for the CESE is based on original work by Ladson-Billings [38] and more recently Mathis & Southerland [39] who maintain that culturally relevant pedagogy, encourages learning by meeting the academic and cultural understandings of diverse students, recognizing and building upon their backgrounds and experiences, and interests in science. Mathis and Southerland [39] point out that scientists often view science as culturally neutral and are uncomfortable merging culture into their classrooms, even though they support beliefs about social justice and equity. These conclusions have been found in their studies of physical science teachers nationwide, using 250 surveys and 25 interviews and intensive case studies [40]. Mathis' instruments for learning more about teachers own understanding of equity and how to obtain reliable measures of science teachers' professional identity and practices will guide us in modifying the CESE science curriculum activities and materials to make it more equitable and culturally responsive to the students' background and their local communities.

I.4 Contribution of the Revised CESE to Communities. Many communities in the south are experiencing; a brain drain (losing residents who are seeking employment in other places), difficulty in recruiting and retaining teachers, limited governmental funding, and disappearance of library, health, and social services [41, 42, 43]. All of which have created a devastating impact on the economic and social well-being of those remaining. However, schools in rural communities are central to the community's vitality including its population stability and growth [44, 45]. Because of the multiple needs of rural schools, the CESE intervention will offer multiple resources including professional learning for the teachers, honorarium for those participating in the study, and curricular materials and science equipment. The team is also committed to raising additional funding, bringing new resources to teachers, administrators, and students with the expectation of supporting the existing communities and encouraging others to offer assistance. One example of is a recent offer by a transportation service to transport teachers, often in remote locations, to where professional learning activities will be held at no-cost. Our longer-term vision is that the CESE intervention will help to inspire more students to attend higher education institutions, particularly in STEM fields bringing new human capital to their families (many of whom have parents with only a high school education) and communities.

II. Strategy to Scale

II.1 Barriers and Strategies to Scale. Four major areas emerged as potential barriers: (1) Revision and Sustainability to the Curriculum; (2) Technological Supplements; (3) Promotion of Access and Equity; and (4) Collaborative work with Local Communities.

Revision and Sustainability to the Curriculum. CESE's main goal is to create an equitable high quality science curriculum for which all students can succeed. Reviewing our materials for scale-up, we became concerned about the responsiveness of the CESE curriculum to the life experiences of adolescents in different locations throughout the US. Although the CESE curriculum is not scripted but intentionally developed to be flexible and adaptable to the classroom community, we had to ask ourselves was it truly responsive to the diverse populations found in distinct often overlooked geographic regions? The issue of responsiveness is especially critical for students in-need, as their science instructional materials are often irrelevant to their lives and communities in which they live [46]. If the goal of our intervention is to reach all students, we needed to modify our teacher and student materials and elevate its capacity to be culturally responsive to different populations, especially for those students most in-need. We deliberately choose the rural south because of our commitment to equity and the secondary science demands of the deep south [47, 48, 49].

The addition of culturally responsive chemistry and physics material to the unit lessons has to be carefully undertaken or it will be ineffective and potentially alienating. To avoid problems of misunderstanding and lack of experience, we sought the advice and collaboration of experts on this topic, including our Co-PI Professor [redacted], a former physics teacher whose research and scholarship are centered on culturally relevant and responsiveness of teachers, and our partners at Alabama A & M and WSSU, who understand contributing factors to the challenges students in the rural south face in science learning. Second, to ensure that the new culturally responsive curriculum materials are valid and reliable, throughout the development period and field test, we will turn to the cooperating field-test districts, administrators, teachers, and students for their input and feedback. Item construct and content will be validated in cognitive labs and interviews, and a series of internal and external reliability tests will be conducted before their inclusion in the efficacy study. Third, that the integrity of the CESE framework be maintained and new material, activities, and assessments remain internally coherent to the learning goals with strong inter-coherence across them [50]. Professor [redacted] [redacted], the Co-PI and creator of the CESE curriculum, will monitor and evaluate all new items ensuring there is strong intra- and inter-coherence across the units and the new culturally responsive items and activities enrich and support the learning goals, instructional material, and skill development of the CESE initial design and framework.

Technological Supplements for Teacher knowledge. Additional barriers central to this work are several technological factors. To promote well-being and science learning, we will modify the current experience sampling method (ESM; [51]), a digital program installed on cellphones that obtain repeated measures on student activities and their feelings about it in real time. The information obtained from the ESM identifies when the students are most engaged, which activities they enjoy the most and least, and how they feel when undertaking new challenges. This will provide teachers with invaluable data on student emotions at random specified times to understand barriers and challenges that may be affecting their emotional well-being. Another potential technological barrier is our use of technology in the classrooms. We will be working with the school districts and local companies like Google to assist us in the purchasing of chrome books for the students, wifi coverage and speed in the schools especially in low-income rural areas. And finally, technology is a critical part of science learning in the CESE curriculum. At all professional learning meetings when involved in the lessons, we will have the teachers participate in technological modeling and other experiments, raising their awareness and expertise for using these tools in their classrooms.

Promoting Access and Equity. As research has shown, many US high schools serving underrepresented minorities have limited materials to conduct “hands-on” science experiments. Given the resource constraints in the South and that materials may be a problem, we have enlisted Four Piping Brook LLC to help provide our schools with the materials including STEMSIMS virtual experiments. Additionally, we will extend and deepen our accommodations for students with learning disabilities. Our subcontractor also has accessibility to digitized material for audio and sight compromised students. With respect to ELLs, we have already translated all material into Spanish and will update them for this study, provide additional translated materials for migrant and other populations that are included in the sample. The majority of high school students with learning disabilities typically need services such as extra time for completing their work and assistance with reading and mathematical computation, which we will provide through technology. Results from the ESM will also help to identify students with disabilities who appear disengaged. We will provide teachers discourse tools to assure that students are not marginalized from participation and active engagement in CESE activities.

To guarantee all states, districts, and schools have access to our materials, we will be placing them as we have in the past, on Creative Commons, an open-source venue for public use. Our partner universities and school districts will also help in the continued and expansion of the dissemination of the CESE intervention for its future sustainability.

Collaborative work of community partnerships. In assembling our team, we were fortunate that several of our colleagues including Professor [redacted] and Professor [redacted], the Department Chair in Educational Leadership, were willing to join our team, both of whom grew up in the south, and share a deep commitment and knowledge in developing a curriculum that is culturally responsive to the lived experience of teachers and students. Our Dean [redacted] [redacted], who also grew up in the South, has assisted us in forming our partnerships with Alabama A&M and WSSU and their expert science educators and researchers and providing substantial no-cost contributions to this grant. Fundamentally, all our interests are in learning more about how todays' teachers identify themselves and their commitments to equitable instruction and the cognitive and social challenges of their students' lives, which make our work not only culturally responsive but also grounded in cultural relevance.

When we conducted our earlier CESE efficacy study, we were able to form a successful partnership with the University of Helsinki, based on the idea that our work was dedicated to advancing the knowledge, skills, and well-being of adolescents. We will follow many of the same activities we engaged in this prior collaboration, including meeting virtually every other week, sharing all instruments and data analyses, conducting seminars on new technologies, reviewing designs, working with each other's graduate students and post-doctorates and co-publishing in journals, giving presentations at meetings, and seeking additional funds for our work. To create such an environment of relational trust, there has to be respect for one another, recognition of each other's competence and expertise, follow-through on commitments, and willingness to place the academic and personal well-being of the students first [52]. These are relationships we will establish with our university partners and participating districts and school administrators and teachers.

II.2. Management plan. A successful project needs the support and willingness of a dedicated team with a shared vision and achievable goals. Showing its support and commitment, MSU, Vice President for Research and the Dean of the College of Education have provided [redacted] [redacted] with 27 percent (3 months effort) released time to lead this project for all five years (see, cost share agreement). [redacted], an experienced director of large grants; in collaboration with her MSU colleagues, [redacted], lead curriculum designer, [redacted], expert in cultural responsiveness, and [redacted], organization specialist in educational leadership; will all be deeply involved in the revisions and implementation of the intervention, collaborative partnership activities, and dissemination and outreach of project products. We are fortunate to have a partnership of exceptional colleagues, Professors [redacted], Interim Department Chairperson, Program Coordinator of Secondary Education and Associate Professor of Science Education at AAMU, and Professor [redacted], Associate Professor of Physics at WSSU, who also will work on intervention revisions, professional learning, and developing teacher leaders for school data collection. The independent evaluator of this project is Professor [redacted]

[redacted], Department of Statistics, Northwestern University and Co-Director of the STEPP center, a widely respected leading expert in RCT models, sample generalizability, and in addition to her own funded projects, has been the statistical consultant on multiple EIR and IES grants. A more extensive explanation of the management plan, including defined responsibilities, timelines, and milestones can be found in Appendix J. Additional partners in formation include various rural LEAs in both North Carolina and Alabama. Because of short timeline, we have two letters, one in each that demonstrate our access and ability to secure cooperation of these LEAs. In North Carolina, we have a prototype MOU with the rural school district name redacted, that we will adapt for the field test. In Alabama, we have a letter of support from the Huntsville superintendent of schools, which forms a hub of collaboration with the surrounding rural school districts. Appendix C provides the prototype MOU and the support agreement with Huntsville.

II.3 Capacity to scale. Much of [redacted] research has examined and shown why we need to scale and how to design and implement evidence based RCTs [53, 54, 55]. To bring an efficacy study to scale, there are several considerations [56]; one of particular relevance here is identifying interventions likely to produce "effects" across different populations and settings. It can be difficult to predict these effects in rural schools as they are geographically diverse and clustered by diverse subpopulations. Our proposal is designed to create a model of making a science intervention culturally responsive, through modification and adaptation for rural school populations in the south, which are often underrepresented in large-scale studies. Forming a partnership with science scholars and educators in the south is essential for creating a reciprocity of ideas and actions for modifying interventions. Professor [redacted] in collaboration with our faculty partners will be constructing a historical narrative of these exchanges, capturing how we adapted CESE to rural students in dissimilar school contexts. This process for ensuring cultural responsiveness to our intervention will complement the rigorous standards for conducting an efficacy trial and subsequent scale-up considerations when extending the CESE to a generalizable population.

II.4 Mechanisms to Disseminate. Three main mechanisms of dissemination will be used in this study: (1) public access to project a) curricular material, professional learning videos, and student unit assessments which are open source and available through Creative Commons; b) project instruments and items available on the CREATE for STEM website; and c) datafiles available in accordance with new IES procedures and MSU policies for distribution of restricted data for replication, reproducibility, and other analyses. (2) Social media presence a) on the study website and links to CREATE for STEM and partner websites which will describe study synopsis, team members' profiles, reports, publications, and other related project information such as presentations and workshops, videos of public events in schools, interviews on radio, TV, and news releases—all with electronic links; b) external media communication for public audiences including Facebook, YouTube, Instagram, Conversations; and (3) scholarly outlets including journals; books; reports; publications for widespread public circulation; and presentations at scientific meetings.

II.5 Utility of Products. All of our research and products are designed to be transparent, replicable and useful. CESE units (including lessons, materials and activities) are accessible on a website/google drive and have been adopted by several districts and countries, China, Finland, South Africa, South Korea and OECD documents [57]. We will soon be adding open access CESE professional learning videos for teachers of in-classroom use.

III. Project Design.

The CESE intervention design uses a system approach composed of PBL units, teacher and student activities and materials, teacher professional learning, and student formative assessments. The foundation of the CESE units is based on six principles formulated by [redacted] [58] which include: 1) Meeting important learning goals, which represents a major shift from being simple content standards to ones where learning goals are supported by explicit scientific practices. 2) Constructing a meaningful driving question, which constitutes the anchoring phenomenon or problem that students attempt to answer throughout the units. 3) Providing opportunities for learners to explore phenomena using scientific practices such as planning and carrying out investigations, analyzing and interpreting data, and constructing explanations and designing solutions. 4) Integrating learning tools to make sense of evidence by incorporating technologies that support inquiry for problems students are likely to encounter in our increasing complex scientific world. 5) Creating collaborative activities that help students find solutions to the driving questions, which allows for rich discussions and exchange of ideas to build knowledge, similar to the work of scientists and engineers who merge many perspectives when testing claims. 6) Using tangible artifacts and assessment tasks that address the three dimensions of learning and capture students' emerging understandings. Artifacts and assessments take multiple forms, and their purpose is to immerse students in the scientific practices, such as constructing and testing system-based models with computer software exposing them to different problems across various contexts.

Science Units. The actual units are created by a team of science education researchers, including NGSS and PBL experts, working with science teachers. The development of each unit was led by a teacher expert responsible for writing the materials with input from team members. Throughout the process, faculty science subject experts, including those who taught undergraduate science courses, also reviewed the units. The team begins with the NGSS process of identifying and unpacking target performance expectations, then using NRC dimensions of learning, articulate disciplinary core ideas, scientific and engineering practices, and crosscutting concepts. Next, relevant driving questions and anchoring phenomena are chosen. Then, a storyline of the unit is developed, including descriptions of each lesson's learning goals expressed as performance expectations, lesson-level driving questions, main activities, and assessment items. In addition, the team creates student packets, teacher guides, and a required material list. This results in seven chemistry units: evaporation, simple chemical reactions, combustion, conservation of matter, equilibrium, nuclear reactions, and energy, and seven units in physics, collision, acceleration, magnetic field, electricity, electric motors, sound waves, and radiation. Unit driving questions, performance expectations, and sequence of learning activities are found in Appendix J.

Post-Unit Assessments. Post-unit assessment tasks were designed to extend student learning experiences by using the three dimensions of scientific knowledge to explain phenomena and solve challenging problems to demonstrate mastery of NGSS performance expectations. The steps for creating these assessment tasks and rubrics are a modification of a previous process articulated by Harris et al. [59]. The development of assessment tasks allowed for the creation of items through a principled, clearly defined process that is grounded in learning and assessment theory. All the post-unit assessments have the students design and explain their own models, which are then evaluated with a rubric that assesses their knowledge of the NGSS performance expectations. A recent analysis of the 2018-2019 CESE study shows that the post-unit assessments have criterion validity to the Michigan state standardized science assessment [60].

Professional Learning. Treatment teachers will receive approximately 30 hours of professional learning where they are introduced and experience what their students will engage with during their science activities, including using the driving question board, building models, developing evidence-based explanations, and conducting experiment. These thirty hours include two days during the summer, prior to the school year as well as two days during the academic year where teaches are introduced to the upcoming units, lesson plans, materials, and activities. In subsequent meetings, teachers are encouraged to share what went well in their classrooms, their challenges, and opportunities for changes in the curriculum experiential activities with the research team and each other. In all this work, technology is used to allow for online collaboration including chats, webinars, document sharing, and exemplar teaching instruction via video [61] that can be monitored for estimating dosage as well as fidelity of implementation.

III.2 Goals of CESE. The teacher and student curricular materials and activities, the unit assessments, and the professional learning address five major short, mid, and long-term goals in secondary science learning. Using project-based learning which has the students figure out challenging questions that are relevant to their lives. The short-term goals are to increase students 1) engagement in science classes and 2) science learning and understanding throughout the school year. The mid-term goals are to increase students 3) science achievement on standardized tests and 4) interest in college and STEM. The long-term goal is to increase students 5) enrollment in postsecondary education and STEM majors. See Logic Model in Appendix G.

III.3 Addressing Needs of the Population. This replication study will treat 4000 (3600 from the efficacy and maturation study and 400 from the field study) new students in their academic and social and emotional learning and 90 teachers where it is expected to find transformation of practices more aligned with PBL principles, NGSS, and the NRC *Framework*. An often-overlooked population, CESE adaptations for students in the rural south will address the specific needs of this population in science learning. At the school, student, and teacher levels we expect the results to: influence a larger proportion of high schools to offer the CESE intervention; students to take chemistry and physics classes especially underrepresented minorities who have been excluded from these courses and science learning; and teachers to implement with fidelity CESE science instruction. We expect that the CESE intervention will raise the importance of science learning so that new generations of students are prepared to face the challenges of the future. Furthermore, with raising additional support, we will begin to address the technological and material needs of this overlooked population.

IV. Project Evaluation

Professor [redacted] and her evaluation team at Northwestern University are the independent evaluators for the CESE Intervention in the South. The evaluator's role includes in periods: 1, planning of the evaluation; 2, field testing data collection and instruments; 3, conducting the test; 4, analyzing the results; and 5, consulting on scale-up. This evaluation uses a block cluster RCT, where treatment is at the school level, to examine the effects of CESE on student academic, social, and emotional outcomes.

Research Questions. (1*) Academic outcomes*: Do students who received the treatment outperform students in the control group on a summative science assessment? Does the treatment effect hold across race and socioeconomic status? Is there a difference in college plans, majors, and STEM career awareness between the treatment and control students? (2) *Social and Emotional outcomes*: Are the treatment students more likely than the control students to find their science classes challenging, interesting and more willing to stay on certain activities and collaborate with classmates? Are the treatment students more likely to be engaged in their science classroom than the control students? (3) *Fidelity of Implementation*: We suspect that variation in teacher instructional practices such as having students engage in specific practices, such as planning investigations, providing explanations, and computerized modeling, will affect the magnitude of the treatment effect. What specific instructional techniques showed the most and least impact on the overall treatment effect? (4) *Cost analysis*: What is the cost benefit of the effect of CESE?

Comparison Condition. The treatment (T) teachers will receive the CESE Intervention, including teacher and student materials, assessments, and professional learning. The Control (C) teachers will carry on with business as usual and will receive a shortened professional learning on NGSS. At the onset of the intervention, the T and C teachers will complete a survey to ensure that they have a similar understanding of NGSS, three dimensions of learning, and PBL. During the year, both T and C teachers will be observed in their classrooms regarding their instructional science practices to confirm that the T teachers are engaged in CESE and the Cs are not. We will also obtain comprehensive information on the C teachers' curriculum activities and student experiences in their classrooms. T and C teachers will also receive an exit survey and ESM surveys to further understand their classroom conditions (e.g., materials, practices).

Study Sample and Randomization

Eligibility and Recruitment. *Schools***.** In each district, traditional public high schools with a minimum chemistry or physic student enrollment of 10 (not including virtual or alternative schools) will be eligible for recruitment. In period one (Jan 2024-Jun 2025), [redacted] will provide the sample of schools from North Carolina and Alabama for the efficacy study. Then, project staff will work in coordination with local university partners and district staff to reach out to schools about the intervention. Initial outreach to school leaders will be made by mail and phone with a follow-up informational packet. Meetings will be scheduled with school principals to explain the intervention and research procedures. Once schools agree to participate, a MOU will be generated outlining procedures for obtaining teacher consent (and student consent where opt-out policies are not in place; see Appendix C for prototype and current MOUs). Facilitators will support the collection of IRB-approved consent forms for teachers and students and answer questions from teachers. This model was used successfully in our prior CESE intervention, and we anticipate being able to recruit the necessary number of schools.

Classrooms. All grade level Chemistry and Physics classes in the schools will be included in the evaluation.

Students. All students who are enrolled in the grade level Chemistry and Physics courses by the end of the first two weeks of school (when most high school schedules are finalized) will be included in the evaluation. Students who join the course later in the year will be excluded from the study sample.

Random Assignment. Randomization will occur at the school level, to lessen the probability of spillover between teachers and classrooms. School lists will be identified from state data, and we will be careful to exclude schools that are scheduled for consolidation or closure in our initial sample. After obtaining the list of schools, we will identify the number of chemistry and physics teachers within each. We then block on number of chemistry and physics teachers to ensure balance of number of eligible teachers between treatment and control schools.

Schools will be randomized within blocks to assure that treatment and control schools are equally represented across students' characteristics such as student race/ethnicity, percent of ELL, percent of students with learning disabilities, average high school science achievement, and percent of students receiving free or reduced lunch to address concerns of differences in program implementation. Since many of these characteristics may be correlated with population outcomes, this will help ensure T and C are represented across the schools, and the randomized block design will improve statistical power.

Randomization. [redacted] will create a dataset where each row represents a school and its associated geographic and demographic characteristic variables are entered in columns, including its assigned strata. Schools will be randomized to treatment and control conditions.

Expected Sample Sizes. The data from our 2018-2019 efficacy trial provide the best evidence to date of the effect that a high school chemistry and physics PBL curriculum intervention can have on students' science learning to be used in a power analysis. Analysis of the efficacy trial data indicates that the CESE treatment had an effect size of .21 standard deviations compared to the control group. This effect size is in line with other PBL trials including Harris and colleagues' (59) cluster-RCT of a middle school PBL curriculum where they found effect sizes in the range of .22 to .25. These effect sizes are also consistent with those observed in meta-analyses of a broader range of science curricula [62].

Using the parameters from the efficacy study (number of students per school = 61, school level ICC = 0.177), we estimate that we would need 49 schools to achieve a 0.8 power for a minimum detectable effect size of 0.20. This is based upon power analysis for cluster RCT that use a hierarchical linear framework. Based upon our efficacy study's 13% school level attrition rate, we would need to recruit at least 56 schools. To ensure an even number of schools between Alabama and North Carolina, 30 schools will be recruited from each with 15 treatment and 15 control schools in each state. Schools will be recruited from school districts in Alabama and North Carolina.

Key Measures and Plan for Obtaining Data. The data collection components and timeline can be found in Appendix J. These data collection activities will require collaboration and organization across research partners and school districts. Partnerships and data collection procedures will be developed and finalized in the first 30 months of the project period and thoroughly discussed and reiterated between the CESE team and their partners.

Data Collection. *Student and Teacher Surveys***.** All surveys and assessments will be self-administered online (through Qualtrics) with the facilitator's supervision. The facilitators who will be trained by our partner universities in North Carolina and Alabama will answer questions, troubleshoot problems with Qualtrics, and encourage completion of surveys by teachers and students. Survey content can be found in Appendix J. The evaluation research team will support facilitators in data collection by assisting with scheduling, providing technical support, and monitoring of survey response rates. These research team will also monitor Qualtrics completion reports on a weekly basis to identify areas of non-response and notify facilitators, who can then contact those sites to answer any questions and encourage sites to complete surveys.

Student Assessments. Students will take a science pretest before the intervention begins and a science summative assessment at the end, administered online in their science with support from the facilitator. In the first two months, the science pretest will be given to allow enough time for finalization of enrollment and schedules. In the last two months of the school year, the summative will be scheduled and given to avoid state or district testing to reduce student burden and fatigue.

Teacher Observations. In Period 1, the teacher observers will be recruited and trained on the observation and data collection protocols. They will work closely with the project manager at Michigan State University and the evaluation team Northwestern University. In period 2, we will analyze their observations to ensure that the measures are reliable in measuring fidelity of implementation which will be used in periods 3 and 4. In the efficacy study (Period 3), three observations will be conducted on the T and C teachers. To maintain comparability, we will ensure that site visits across districts occur in a similar time frame. During the maturation study, observers will repeat observation data collection with the C schools, now receiving treatment.

Student ESM. Facilitators will work with teachers and students in their assigned schools to encourage student downloads of the PACO (Personal Analytics Companion) smartphone app for future ESM data collection. PACO is a free, open-source software designed by Google engineer Robert Evans (www.pacoapp.com), which allows us to schedule and randomize ESM prompts to illicit repeated measures in the intervention periods.

Scheduling for the ESM data collection will occur during the field test, the full efficacy study, and the maturation study. During each of these periods, student will have three different ESM data collection periods of three days each. These guidelines may differ depending on the schedules at specific schools (e.g., block schedules that do not meet every day). ESM will be scheduled to signal the student to respond eight times during the day between 7:00 AM and 7:00 PM (similar to previous ESM studies such as Csikszentimihalyi & Schneider, [63]) Of the eight signals, three will be scheduled to go off in the student's science class while the other five will randomly signal throughout the rest of the day. Capturing responses outside of science class and outside of the school day will allow for comparisons of student experiences in science, school, and non-school contexts. The ESM questionnaire has a core battery of questions administered each time with additional science-specific questions administered when a student is in science class. During science class, both students and teachers will be signaled at the same time to minimize disruptions throughout the hour and to give us better information about what is happening during several specific moments during the lesson. We will not have the students' data collection period extend through the weekend. The Q-statistic of 2018-19 ESM data shows significantly between-person variance, and the estimated reliabilities varied between different indices, with reliability coefficients exceeding .95 and coefficients ranging between approximately .70 and .90 for intraindividual standard deviation (ISD).

Administrative Data. Administrative state data for teachers and students (e.g., prior achievement and attendance data) will be obtained and used for sampling frames and control variables in analyses. Both MSU and partner universities will work directly with districts to set up Data Sharing Agreements in period 1. Cost analysis data will come from administrative data records of costs related to intervention materials, personnel, and additional costs.

Analytic Measures. The analytic measures include outcome, baseline, independent, and mediating variables. Additional information on these can be found in Appendix J.

Outcome measures. The primary outcome measure is student science achievement. Auxiliary outcomes include two additional measures, one on student social and emotional learning, and the other on student STEM college and career ambitions. These will be measured from the summative science assessment, the ESM, and the student exit survey. The science achievement is measured using a summative assessment developed by the Michigan Department of Education (MDE) to measure student science proficiency on the corresponding NGSS performance expectations for chemistry and physics. To make the appropriate comparison between students who took the assessment in different subjects, this study conducts a raw score transformation equating and standardized procedure to obtain a comparable z-score. To determine the reliability of the summative assessment, we first ran an exploratory factor analysis where we found 3 factors: 1 physics and 2 chemistry. Since the items are dichotomous, we used a KR-20 reliability test. The reliabilities are: 1) for Physics, 0.69 2) for first chemistry factor, 0.60, and the second, 0.45. Student social and emotional learning is measured from items on the ESM survey. In prior work, we have used the ESM to examine affective measures such as challenge, skill, interest, enjoyment, concentration, the importance of living up to one's expectations and those of others, imagination, and problem solving. (Information on the reliability and validity of the ESM can be traced to the 1980s and subsequent use and their reliabilities can be found in previous studies [64, 65, 66]). The reliabilities of these measures will be verified in period 2. Finally, the measure of student STEM college and career ambitions comes from two items on the student exit survey that asks about plans for post-secondary education and career interests.

Baseline measures. Baseline equivalence will be assessed on the T and C students using another broad science pretest, social and emotional learning measures, student college and career ambition measures from the student background survey. This pre-test will contain multiple choice and several constructed response items, chosen from the National Assessment of Educational Progress (NAEP) test bank and several items aligned with disciplinary core ideas and performance expectations commonly taught in chemistry and physics classes. The quality of the pretest was verified using a multinomial logistic regression (MLR) and an item response theory (IRT) nominal response model, indicating similar response pattern for students in chemistry and physics. The assessment reliability is about .90. The background survey is the exact same survey as the exit survey but administered at the beginning of the school year. Finally, using administrative data on the students, baseline equivalence will be also assessed on student demographic information, including socioeconomic status, using whether they qualified for free and reduced lunch, race and ethnicity, English language learner status, and IEP status.

Independent variables. The independent variables for analysis of the treatment effect will include the treatment indicator and the randomization block.

Mediating variables. For the analysis of mediation effects, two measures will be used. The first will be teacher practices reported from teacher observations. The protocol for the teacher observations includes 26 measures of teacher and student practices and activities in the classroom. The reliability of teacher implementation in PBL classroom observation is 0.85 whereas the student engagement in PBL classroom observation is 0.83. An additional measure for the mediation will come from the student ESM data from the item where the student responds with what activity they are participating in their class.

The teacher survey included perceived role in teaching, familiarity with PBL, teaching activities, mindset, burnout, job satisfaction, and PD participation. The reliability of the constructs ranged from 0.77 to 0.91, indicating moderate to high reliability in the survey questionnaire. We also identified several significant correlations between classroom-observed teaching behavior, teacher perception, mindset, and PD. For example, teachers perceived to have high support in inquiry-based teaching, reasoning, and thinking are significantly associated with more teacher discourse moves and small group discussions in the classroom.

Analysis of Treatment Effect Model. To answer Research Question 1 regarding the intervention's impact on student science learning outcomes, we will use hierarchical linear models (HLM), specified as student-in-school with a binary treatment condition at level 2. In the efficacy study, we found little variation by teacher; therefore, a two-level HLM was deemed appropriate.

Equation (1): The impact of treatment on learning in science

$$Y_{ij} = \beta_{00} + \beta_{01} T_j + \beta_{0j} X_j + \beta_{03} B_j + r_{0j} + \epsilon_{ij}$$

Y_{ij} is the person-level achievement outcome for person i in school j. ϵ_{ij} is the person-specific error term. β_{00} is the mean achievement outcome. β_{01} is the treatment effect. T_j is a binary treatment indicator for school j. X_j is a vector of school-level covariates, including school-aggregate science scores and demographic variables. B_j is the fixed effect for the strata used in the block randomization. r_{0j} is the random effect associated with each school. The outcome here is the summative assessment and the variable of interest is the treatment indicator at the school level.

Subgroup Analyses. To answer the additional question regarding whether there is a treatment effect across race and socio-economic status, the above treatment effect model will be estimated for different subgroups (race/ethnicity, free/reduced lunch).

Exploratory Treatment Effect on Secondary Outcomes. For the research question regarding the intervention's effect on college plans and STEM awareness, we will use a series of generalized linear mixed models (GLMMs). For Research Question 2 regarding the treatment effect on student's social and emotional learning, we will also use a series of GLMMs with the outcome of social and emotional learning variables, such as finding their science classes challenging, interesting, and willingness to stay on task. Additionally, we will use a structural equation model to understand the relationship between the treatment and the variables corresponding to the construct of engagement.

Mediator Analyses. For Research Question 3 regarding the fidelity of implementation of the intervention, we will use exploratory mediation analysis. Combining ESM results on students' social and emotional experiences and reports of what types of activities they are engaged in with teacher ESM, videos, and observations will shed light on the mediators that explain estimated effects. In addition to looking at ESM results as a mediator for treatment effects, we will also use a variety of HLM and SEM models to investigate the situational variables that impact student classroom experience.

The mediation model for fidelity of implementation is given in Equation (3):

Equation (3): The upper-level (3-2-1) mediating impact of fidelity of implementation on science learning

$$M_{jk} = \gamma_{000} + \gamma_{001} T_k + \gamma_{002} B_k + r_{00j} + \upsilon_{0jk}$$

$$Y_{ijk} = \grave{\gamma}_{000} + \grave{\gamma}_{010} M_{jk} + \grave{\gamma}_{002} T_k + \grave{\gamma}_{003} B_k + \grave{r}_{00j} + \grave{\upsilon}_{0jk} + \grave{\epsilon}_{ijk}$$

Where:

M_{jk} is a fidelity of implementation measure for teacher *j* in school k. T_k is a binary treatment indicator for school *k*. Y_{ijk} is science learning, as demonstrated by performance on the summative assessment for student *i* with teacher *j* in school k. The mediating effect of fidelity of implementation is calculated as: $\gamma_{001} * \grave{\gamma}_{010}$. Again, we can test the significance of the resulting mediation effect using BC parametric bootstrap.

Handling Missing Data. Students with missing outcome data will be excluded from the analysis. For students are missing covariates, dummy variable adjustments will be used.

Joiner Considerations. This blocked cluster RCT presents a low risk of bias due to individuals entering clusters because school treatment status will not be known to parents and students outside of the study.

Attrition Considerations. We have found that the support of districts, school leaders, and teachers, participation in the study is positive and few schools or teachers have chosen to attrit. However, there are circumstances beyond our control that may lead schools and teachers to attrit from the study, such as strikes, consolidation procedures, and pandemics. The power analysis for this proposed study uses conservative estimates of all relevant parameters that are informed by our prior work, so that our sample will be of adequate size to absorb some attrition. Combined with our plans to ensure baseline equivalence (see below) and representation of the analytic sample to the clusters, this study should be eligible to meet WWC v 5.0 [67] standards without reservations.

Baseline Equivalence Testing. Because this is a RCT, any imbalance between the control and treatment students should be due to chance. However, in the case of differential attrition between the T and C conditions, baseline equivalence of the two samples will be assessed using the analytic sample (students who are not missing the outcome variables). This will be evaluated for each analytic sample (if the analytic samples differ by outcome variables). To assess the baseline equivalence, we will estimate a similar model to equation 1 above, with the baseline variable as the outcome variable in the two level HLM and the predictor of the treatment indicator and randomization block. The estimate of the coefficient on the treatment indicator will be our estimate for the magnitude of the mean difference between T and C students. This estimate will be divided by the pooled standard deviation to get the standardized difference (Hedges' g). If this standardized difference is < 0.05, it will be considered equivalent. if the difference is between 0.05 and 0.25, the variable will be included as a covariate in the analytic model.

IV.1 What Works Clearinghouse. To meet the *WWC Standards v. 5.0* [67], this project addresses the four major considerations from the standards: outcome measures and checking for confounding factors, assignment to conditions, compositional change, and baseline equivalence. The section above on analytic measures reports the outcome measures, their face validity, reliability, and addresses that these measures are not over aligned with the intervention. In the data collection, these measures are also consistent across the sample and thus meet the requirements for the outcome measures. With regards to confounding factors, the randomization process will ensure that there is no additional confounding variable in the T condition. Assignment is reported in the randomization section and follows the standards for random assignment using block randomization and maintaining the integrity of the assignment. All units included will be randomly assigned and the analysis will be conducted using Intent to Treat. No analyses will exclude units based on reasons related to the treatment status. Attrition will be measured and evaluated at the cluster and individual level. The above section on joiners and attrition addresses meeting the standards for compositional change and ensuring analytic representation of the original sample. Baseline characteristics for will be established for all outcomes using measures that are either the same measure or in the same domain or broader, see section on baseline measures. For every analysis, baseline will be evaluating using the respective analytic sample and adjusted for the blocks from the randomization. From outcome measures to baseline equivalence, this study should be eligible to meet WWC standards without reservations.

IV.2 Acceptable level of implementation. To evaluate an acceptable level of implementation, attendance at the teacher professional learning will be collected and the teacher exit survey for the T teachers will include questions regarding which CESE units were taught. An acceptable level of implementation will be at minimum attending professional learning and teaching at least one CESE unit. Regardless of this acceptable level of implementation, all T and C teachers and students will be in the full sample for the intent to treat analysis.

IV.3 Replication. We will have an electronic codebook for data access protocols for those interested in reproducing the work. Additionally, all instruments and data will be available (following new IES procedures) for researchers interested in replicating and extending our treatment effects to other student populations to expand our understanding of the impact of the intervention in multiple contexts.

IV.4 Cost-Effectiveness of the Intervention. For the one-year implementation of CESE, we will estimate costs using the Ingredients Method [68,69].

Identifying the Resources and Estimating Costs. We will identify the value/price, quantity, and percentage of use for each resource needed to implement CESE, including any resources provided in-kind. We will use national average prices to determine the total cost of implementing CESE. When national prices are not publicly available, we will use cost assessment tools, surveys, and interview protocols to collect these data and adjust the price using geographic indices to covert the local to national prices. Data collected with interviews and surveys will be used to the measure the quantity and percentage of use for each resource. In addition to the total cost, we will estimate additional costs incurred after accounting for the costs associated with business as usual [70]. Because we examine the effects of CESE at the school level, we intend to analyze and report the program costs per school and will also explore costs by school size and geographic location as a sensitivity analysis.

Cost Effectiveness Analysis. To measure primary outcomes and program costs, we will derive a cost-effectiveness ratio, which allows stakeholders to understand the resources needed to produce the observed effects for CESE. However, it may not be possible to derive a comparison of cost-effectiveness ratios. In this case, we will try to identify a comparable alternate program that targets similar outcomes and estimates the costs for implementing this alternate program. From this, we will derive a cost-effectiveness ratio between this alternative program and CESE. If an alternative program cannot be found, we will use existing literature to identify public data on similar ratios from comparable interventions to estimate the relative cost-effectiveness of CESE.

References

[1] Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Klager, C., Bradford, L., Chen, I., Baker, Q., Touitou, I., Peek-Brown, D., Marias Dezendorf, R., Maestrales, S., & Bartz, K. (2022). Improving science achievement—Is it possible? Evaluating the efficacy of a high school chemistry and physics project-based learning intervention. *Educational Researcher*, *51*(2), 109-121.

[2] Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253.

[3] Leung, M., Cardichon, J., Scott, C., & Darling-Hammond,L. (2020). *Inequitable opportunity to learn: Access to advanced mathematics and science courses*. Learning Policy Institute.

[4] Krajcik, J., & Shin, N. (2014). Project-based learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 275–297). Cambridge University Press.

[5] Krajcik, J., & Czerniak, C. (2018). *Teaching science in elementary and middle school: A project -based approach* (5th ed.). Taylor & Francis.

[6] Krajcik, J., Schneider, B., Miller, E., Chen, I., Bradford, L., Bartz, K., Baker, Q., Palinscar, A., Peek-Brown, D., & Codere, S. (2020). *Assessing the effect of of project-based learning on science learning in elementary schools* (technical report, George Lucas Foundation, San Rafael, California.

[7] Krajcik, J. S., Miller, E., & Chen, I. (2021). Using project-based learning to leverage culturally relevant pedagogy for sense making of science in urban elementary classrooms. *The international handbook of research on multicultural science education. Springer*.

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

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

[10] Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice, 38(2), 53–67.* https://doi.org/10.1111/emip.12253

[11] National Academies of Sciences. (2021). Science and engineering in preschool through elementary grades: The brilliance of children and strengths of educators. National Academies Press. https://doi.org/10.17226/26215

[12] Bielik, T., Stephens, L., McIntyre, C., Damelin, D., & Krajcik, J. S. (2022). Supporting student system modelling practice through curriculum and technology design. *Journal of Science Education and Technology*, *31*(2), 217-231.

[13] He, P., Chen, I. Touitou, I., Bartz, K., Schneider, B., & Krajcik, J. (2023). Predicting Student Science Achievement Using Post-Unit Assessment Performances in a Coherent High School Chemistry Project-Based Learning System. *Journal of Research in Science.*

[14] Krajcik, J., McNeill, K. L., & Reiser, B. J. (2008). Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science education*, *92*(1), 1-32.

[15] Shin, N., Bowers, J., Krajcik, J., & Damelin, D. (2021). Promoting computational thinking through project-based learning. *Disciplinary and Interdisciplinary Science Education Research*, *3*(1), 1-15.

[16] Schneider, B., Chen, I.-C., Bradford, L., & Bartz, K. (2022). Intervention initiatives to raise young people's interest and participation in STEM. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.960327

[17] Showalter, D., Hartman, S. L., Johnson, J., & Klein, B. (2019). Why Rural Matters 2018-2019: The Time Is Now. A Report of the Rural School and Community Trust. *Rural School and Community Trust*.
[18] National Academies of Sciences, Engineering, and Medicine. (2019). *Minority serving institutions: America's underutilized resource for strengthening the STEM workforce*. National Academies Press.
[19] National Academies of Sciences, Engineering, and Medicine. (2019). *Science and engineering for grades 6-12: Investigation and design at the center*. National Academies Press
[20] National Assessment of Educational Progress. (2021). *Results from the 2019 Science Assessment.* U.S. Department of Education and the Institute of Education Sciences
[21] Bailey, D. H., Duncan, G. J., Murnane, R. J., & Au Yeung, N. (2021). Achievement Gaps in the Wake of COVID-19. *Educational Researcher*, 0013189X211011237.
[22] National Academies of Sciences. (2021). Science and engineering in preschool through elementary grades: The brilliance of children and strengths of educators. National Academies Press. https://doi.org/10.17226/26215
[23] Showalter, D., Hartman, S. L., Johnson, J., & Klein, B. (2019). Why Rural Matters 2018-2019: The Time Is Now. A Report of the Rural School and Community Trust. *Rural School and Community Trust*.
[24] Schneider, B., Chen, I.-C., Bradford, L., & Bartz, K. (2022). Intervention initiatives to raise young people’s interest and participation in STEM. *Frontiers in Psychology, 13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.960327
[25] He, P., Chen, I. Touitou, I., Bartz, K., Schneider, B., & Krajcik, J. (2023). Predicting Student Science Achievement Using Post-Unit Assessment Performances in a Coherent High School Chemistry Project-Based Learning System. *Journal of Research in Science.*
[26] Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Klager, C., Bradford, L., Chen, I., Baker, Q., Touitou, I., Peek-Brown, D., Marias Dezendorf, R., Maestrales, S., & Bartz, K. (2022). Improving science achievement—Is it possible? Evaluating the efficacy of a high school chemistry and physics project-based learning intervention. *Educational Researcher*, *51*(2), 109-121.
[27] National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. National Academies Press.
[28] National Research Council. (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. National Academies Press.
[29] National Academies of Sciences, Engineering, and Medicine. (2019). *Science and engineering for grades 6-12: Investigation and design at the center*. National Academies Press
[30] National Academies of Sciences. (2021). Science and engineering in preschool through elementary grades: The brilliance of children and strengths of educators. National Academies Press. https://doi.org/10.17226/26215
[31] National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas.* National Academies Press.
[32] NGSS Lead States. (2013). *Next generation science standards: For states, by states.* National Academies Press.
[33] Krajcik, J., & Shin, N. (2014). Project-based learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 275–297). Cambridge University Press.
[34] Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Klager, C., Bradford, L., Chen, I., Baker, Q., Touitou, I., Peek-Brown, D., Marias Dezendorf, R., Maestrales, S., & Bartz,

K. (2022). Improving science achievement—Is it possible? Evaluating the efficacy of a high school chemistry and physics project-based learning intervention. *Educational Researcher*, *51*(2), 109-121.
[35] Showalter, D., Hartman, S. L., Johnson, J., & Klein, B. (2019). Why Rural Matters 2018-2019: The Time Is Now. A Report of the Rural School and Community Trust. *Rural School and Community Trust*.
[36] National Academies of Sciences, Engineering, and Medicine. (2019). *Science and engineering for grades 6-12: Investigation and design at the center*. National Academies Press
[37] Krajcik, J. S., Miller, E., & Chen, I. (2021). Using project-based learning to leverage culturally relevant pedagogy for sense making of science in urban elementary classrooms. *The international handbook of research on multicultural science education. Springer*.
[38] Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American educational research journal*, *32*(3), 465-491.
[39] Mathis, C., & Southerland, S. (2022). Our shifting understandings of culturally relevant pedagogy in physics. *The Physics Teacher*, *60*(4), 260-265.
[40] Mathis, C., Southerland, S. A., & Burgess, T. (2023). Physics teachers' dispositions related to culturally relevant pedagogy. International Journal of Science Education, 1-20.
[41] O'Neal, L., & Perkins, A. (2021). Rural exclusion from science and academia. *Trends in microbiology*, *29*(11), 953-956.
[42] Harris, R. S., & Hodges, C. B. (2018). STEM education in rural schools: Implications of untapped potential. *National Youth-At-Risk Journal*, *3*(1), 3-12.
[43] Johnson, A., Kuhfeld, M., & Soland, J. (2021). The forgotten 20%: Achievement and growth in rural schools across the nation. AERA Open, 7, 23328584211052046.
[44] Sipple, J. W., Francis, J. D., & Fiduccia, P. C. (2019). Exploring the gradient: The economic benefits of 'nearby'schools on rural communities. *Journal of Rural Studies*, *68*, 251-263.
[45] Chetty, R. (2016). Improving opportunities for economic mobility: New evidence and policy lessons. Economic Mobility Research and Ideas on Strengthening Families Communities the Economy, edited by Brown, Alexandra, Buchholz, David, Davis, Daniel, and Gonzalez, Arturo, 35-42.
[46] National Academies of Sciences, Engineering, and Medicine. (2019). *Science and engineering for grades 6-12: Investigation and design at the center*. National Academies Press
[47] Showalter, D., Hartman, S. L., Johnson, J., & Klein, B. (2019). Why Rural Matters 2018-2019: The Time Is Now. A Report of the Rural School and Community Trust. *Rural School and Community Trust*.
[48] National Academies of Sciences, Engineering, and Medicine. (2019). *Science and engineering for grades 6-12: Investigation and design at the center*. National Academies Press
[49] National Academies of Sciences. (2021). Science and engineering in preschool through elementary grades: The brilliance of children and strengths of educators. National Academies Press. https://doi.org/10.17226/26215
[50] Fortus, D., & Krajcik, J. (2012). Curriculum coherence and learning progressions. In B. Fraser, K. Tobin, & C. McRobbie (Eds.), *Second international handbook of science education* (pp. 783–798). Springer.
[51] Schneider, B., Chen, I.-C., Bradford, L., & Bartz, K. (2022). Intervention initiatives to raise young people's interest and participation in STEM. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.960327

[52] Bryk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. Russell Sage Foundation

[53] Schneider, B. L., & McDonald, S. K. (Eds.). (2007). *Scale-up in education: Issues in practice* (Vol. 2). Rowman & Littlefield.

[54] Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., Shavelson, R. J., & White, O. D. A. T. T. (2007). Estimating causal effects using experimental and observations designs. *Report from the Governing Board of the American Educational Research Association Grants Program. Washington, DC: American Educational Research Association*.

[55] Schneider, B. (2022). *Observing Equity and Inclusion in Effectiveness Research*. Hedges Lecture, Society for Educational Effectiveness.

[56] Milat, Andrew, Karen Lee, Kathleen Conte, Anne Grunseit, Luke Wolfenden, Femke Van Nassau, Neil Orr, Padmaja Sreeram, & Adrian Bauman. (2020). Intervention Scalability Assessment Tool: A decision support tool for health policy makers and implementers. *Health research policy and systems*, *18*, 1-17.

[57] Vincent-Lancrin, S., Gonzales-Sancho, C., Bouckaret, M., Luca, F., Fernandez-Barrerra, M., Jacotin, G., Urgel, J., & Vidal, Q. (2019). Fostering students' creativity and critical thinking: What it means in school. OECD Publishing. https://doi.org/ 10.1787/62212c37-en

[58] Krajcik, J., & Shin, N. (2014). Project-based learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 275–297). Cambridge University Press.

[59] Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice, 38(2), 53–67.* https://doi.org/10.1111/emip.12253

[60] He, P., Chen, I. Touitou, I., Bartz, K., Schneider, B., & Krajcik, J. (2023). Predicting Student Science Achievement Using Post-Unit Assessment Performances in a Coherent High School Chemistry Project-Based Learning System. *Journal of Research in Science.*

[61] Krajcik, J., & Shin, N. (2014). Project-based learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 275–297). Cambridge University Press.

[62] Cheung, A., Slavin, R. E., Kim, E., & Lake, C. (2016). Effective secondary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching*, 54(1), 58–81. https://doi.org/10.1002/tea.21338

[63] Csikszentmihalyi, M., & Schneider, B. (2001). *Becoming adult: How teenagers prepare for the world of work*. Basic Books.

[64] Schneider, B. L., & Stevenson, D. (1999). *The ambitious generation: America's teenagers, motivated but directionless.* Yale University Press.

[65] Csikszentmihalyi, M., & Schneider, B. (2001). *Becoming adult: How teenagers prepare for the world of work*. Basic Books.

[66] Schneider, B., & Waite, L. (2006). Timely and timeless: Working parents and their children. In *Work, family, health, and well-being* (pp. 77-89). Routledge.

[67] What Works Clearinghouse. (2022). W*hat Works Clearinghouse™: Standards handbook, version 5.0.* Institute of Education Sciences, U.S. Department of Education.

[68] Levin, H.M. (1975). Cost-effectiveness analysis in evaluation research. In M. Guttentag & E.L. Struening (Eds.), *Handbook of evaluation research* (Vol.; pp. 346–368). Beverly Hills, CA: Sage.

[69] Levin, H.M. (2001). Waiting for Godot: Cost-effectiveness analysis in education. In R. Light (Ed.), *New directions for evaluation* (pp. 55–68). San Francisco, CA: Jossey-Bass.

[70] Hollands, F. M., Pratt-Williams, J., & Shand, R. (2020). Cost analysis standards & guidelines 1.0.