Table of Contents

Project Title: Science Roots: Growing Student Futures through GenAI Enhanced Project-Based Learning Using Green Ninja

Goal - The *Science Roots* project aims to elevate the test scores of high-need students by 10% through enhancing and scaling the Green Ninja project-based learning (PBL) curriculum using generative artificial intelligence (GenAI) tools and innovative teacher approaches.

A. Significance

Science literacy in the US lags behind other countries, and underserved populations fall even shorter (OECD, 2022). Low test scores partly stem from 25% of US middle school students reporting low interest and enjoyment in science classes (NAEP, 2019). Despite strong data showing that engaging strategies like project-based learning (PBL) are highly effective, legacy textbooks that promote didactic, teacher-centered instruction continue to dominate the curriculum marketplace. Low-income and minority students are even less likely to have access to inquiry-based science pedagogies like PBL (Haberman, 1991; Change the Equation, 2017; Gorski & Landsman, 2023; Songer et al., 2002; Thadani et al., 2010). This mid-phase study, the *Science Roots* project, explores how a promising PBL curriculum for middle school science can be enhanced with generative artificial intelligence (GenAI) tools and innovative teaching methods to serve a range of schools including those with underserved populations.

Absolute Priority 1 – Moderate Evidence. This project implements Green Ninja, an innovative middle school science program designed around PBL principles. A What Works Clearinghouse (WWC) report with moderate evidence base showed the positive impact of PBL ([redacted]), noting that PBL requires a driving question that "motivates a solution to a complex problem" and "is meaningful to students' lives." Another WWC report with strong evidence base noted the need for effective technology to deliver high-quality hands-on,

inquiry-based instruction, which was found to promote student use of higher-order thinking skills (Newman et al., 2012). This framework also echoes the recommendations of the moderate evidence base in a WWC Practice Guide that urges teachers to "embed…science activities in interesting contexts" by "using project-based learning…to stir interest in a topic" (IES, 2007a).

Absolute Priority 3 – Promoting Equity in STEM. Researchers have documented a consistent pattern showing that women and minorities value helping others when it comes to choosing careers (Barrington & Duffy, 2007; Corbett & Hill, 2015; Diekman et al., 2011; Hokanson et al., 2007; Meadows & Jarema, 2006; Seymour & Hewitt, 1997). Historically, science curricula have not emphasized these altruistic applications, which may be one of the many reasons why these groups are underrepresented in STEM. When science curriculum actively engages students in solving problems in the community, it especially motivates these underrepresented groups (Boucher et al., 2017; Corbett & Hill, 2015; SciGirls, 2019). Promoting social relevance enhances students' sense of the value of science, which we know influences interest and career choices, especially among women and minorities (Anderson & Ward, 2014; Lewis & Connell, 2005). The Green Ninja curriculum uses this approach to engage students; every project centers around solving a local environmental problem. The *Science Roots* project will focus on low-income, diverse learners who are performing below proficient on the state science test. These students stand to benefit significantly from PBL.

Competitive Preference Priority 1: Implementers and Partners. The university leading this effort is San José State University (SJSU), a Hispanic-serving institutions (HSIs) within the California State University system (CSU). The CSU hosts an initiative, STEM-NET, promoting collaboration across its 23 campuses to make it a worldwide leader in increasing the pipeline of diverse STEM students (California State University, n.d.). Approximately 55% of SJSU students

are Latino/a/x, following in the footsteps of PI [redacted] who graduated with a bachelor's and master's from California State University, Northridge (CSUN) and is currently one of more than 20% of SJSU faculty that identify as Hispanic.

A.1 A problem and an opportunity

Across the US, 42% percent of eighth graders report that they almost never participated in scientific inquiry-related classroom activities (NAEP, 2019). What's worse, high-poverty schools are one third less likely to do inquiry than high-income schools (Change the Equation, 2017). It's more than a lack of science equipment – it's a lack of science pedagogy. With 46% of students spending most of their time reading out of textbooks, it's not surprising that so many students report being disengaged with science (NAEP, 2019).

To activate the 25% of students who report low engagement and to better engage those who are already excited about science, we need to find a way to connect science to things in their world – things that matter to their lives and their futures. PBL engages students in real-world problems and increases student engagement (Carrabba & Farmer, 2018; Condliffe, 2017; Leggett & Harrington, 2021). Despite the promise of PBL, Green Ninja is the only commercially available middle school science curriculum that comprehensively embraces all the key elements of PBL, including the use of real-world problems and giving students a voice and choice in their projects.

A.2 About Green Ninja

Green Ninja is a comprehensive middle school science program that teaches all the Next Generation Science Standards (NGSS) through the context of solving environmental problems. Green Ninja's notable enhancement over traditional PBL is that each project has a broader purpose of improving the home, school, or local community, providing a strong context for

student learning. In the proposed research, Green Ninja's PBL and focus on environmental problem-solving will be further enhanced through a GenAI Co-Pilot to make teaching PBL easier and more effective for teachers. This progression is illustrated in Figure 1, which shows the transition from prior evidence supporting PBL, through Green Ninja's existing success with PBL and environmental problem-solving, to the current project addressing barriers to scaling by integrating the GenAI Co-Pilot.

Figure 1. Timeline of innovations showing how Green Ninja builds on prior evidence.

Green Ninja has been adopted by California and Oregon as an approved curriculum and was one of only three middle school curricula to earn perfect scores on the Texas Education Agency TRR rubric in Texas. Currently, about 20 school districts have adopted Green Ninja as their science curriculum, with about 17,000 middle school students using it daily.

A.3 Green Ninja is Effective

How well do districts using Green Ninja perform? In a quasi-experimental design analysis, we tracked growth on statewide test scores over 4 years for districts that adopted Green Ninja in 2019 and 2020 (**6 districts, ~4,500 students**) in comparison to stratified cohorts of districts that used other curricula but have similar demographics and starting scores (103 districts, ~53,000 students) (Figure 2). The mean scale score of districts in the Green Ninja intervention grew by 2.6 points ($p=0.03$, $d=0.15$) while the controls remained statistically unchanged. The effect is robust across different measures of growth, different stratification parameters, and different demographic groups.

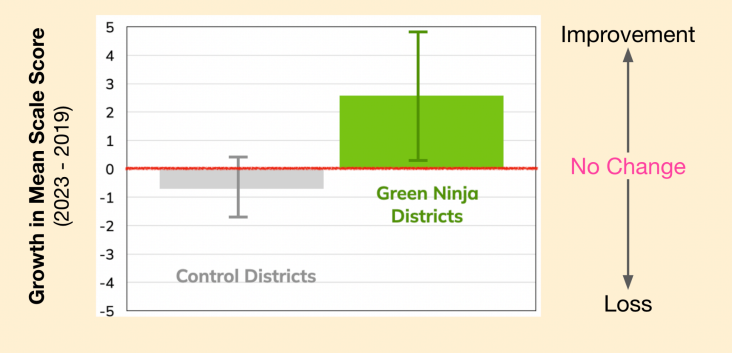


Figure 2. Green Ninja districts outperformed control districts on statewide tests.

Green Ninja is effective because the PBL "unit challenges" provide a **purpose** for learning and **empower** students to solve real-world problems. Together, these lead to stronger **engagement** with the content and ultimately deeper **learning** (Figure 3). The purpose comes from having a driving problem for every unit (IES, 2007a). The empowerment comes as students successfully propose and implement solutions to these problems. While we use these everyday terms of purpose and empowerment, they correspond to *utility value* and *outcome expectancy* in the theoretical framework of Situated Expectancy-Value Theory (SEVT) (Eccles & Wigfield, 2002; Eccles & Wigfield, 2020) where robust evidence supports the causal links in Figure 3, including both short-term learning and long-term career interests (Ainley & Ainley, 2011; Hulleman & Harackiewicz, 2009; Maltese & Tai, 2011; Updegraff et al., 1996).

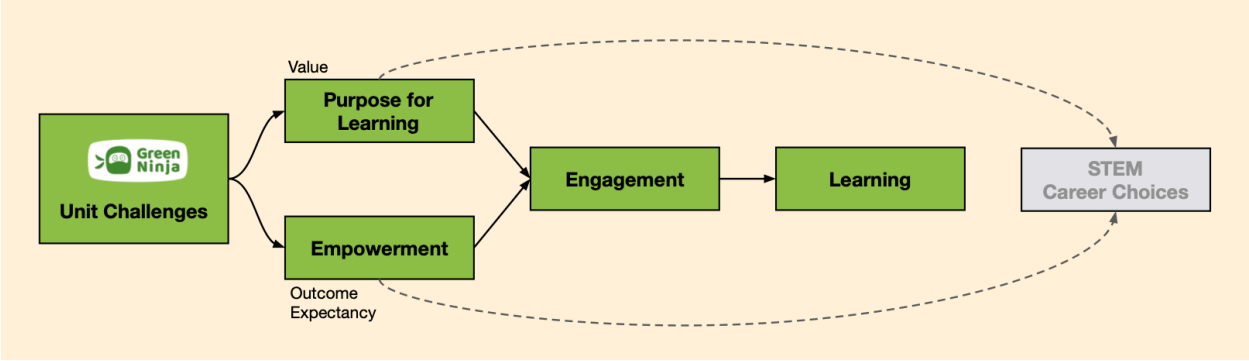


Figure 3. Green Ninja's theory of change based on situated expectancy-value theory.

The few existing studies that have examined PBL's influence on value have focused on how students assess the value of the project to their individual future (agentic or personal value) (Cooper, 2014; Cooper & Kotys-Schwartz, 2022; McLean et al., 2019), but we postulate that Green Ninja's focus on the environment may also enhance students' *communal value* – the sense that STEM is useful to help the broader community. As discussed above in Absolute Priority 3, we know that communal value is a driver for underrepresented populations.

In Green Ninja, students also gain agency as environmental problem solvers. Table 1's case study illustrates how students successfully implement their solutions, leading to personal success that enhances their outcome expectancy for using STEM to solve future problems.

Table 1. Case Study: Green Ninja Transportation Unit.

Unit Challenge	Propose ways to make your community's transportation system more efficient.
Phenomenon	Different vehicles use different amounts of energy to go the same distance.
Science Concepts	As students compare the energy used by bicycles, small cars, loaded trucks, maglev trains, hovercrafts, helicopters, and bulldozers, they discover physical science concepts of forces, motion, magnetism, and kinetic energy.
One Student's Solution	Bella understands that the amount of energy it takes to move a vehicle depends on its mass. She realized that her family had lots of extra stuff in their car's trunk – extra mass that made them waste energy. She proposed the "Junk in the Trunk" clean-up day and organized her entire class to get their parents to participate. While the energy savings were relatively small, she felt empowered knowing that she could put science knowledge into action and organize her classmates to scale her solution.

A.4 The Innovation - Green Ninja Co-Pilot

Green Ninja's PBL would be even more powerful if it were easier for teachers to manage and if there were stronger scaffolding to help students link their projects to science learning. This is especially important in schools serving high-needs students, where teachers often face greater challenges and may have less preparation for innovative teaching methods. With this proposal, we will harness the revolutionary potential of Generative AI (GenAI) technology to develop the Green Ninja Co-Pilot (GN Co-Pilot). The GN Co-Pilot will address the perception that PBL is

too time-consuming or challenging by helping teachers develop personalized assessments, grade student work, and provide real-time feedback to students in a matter of seconds. It will support student learning, particularly as students develop their projects and then connect these projects to science learning. By automating formative assessment and offering personalized support, the GN Co-Pilot aims to make PBL more accessible, manageable, and effective, significantly increasing the time-to-value ratio.

The Green Ninja team has already introduced some basic GenAI system components (Figure 4), and a growing body of literature provides ample evidence of the potential value of GenAI systems within educational environments (Lee & Zhai, 2024; Bahroun et al., 2023; Cooper, 2023) and the need to ensure ethical applications (Alier, et al., 2024).

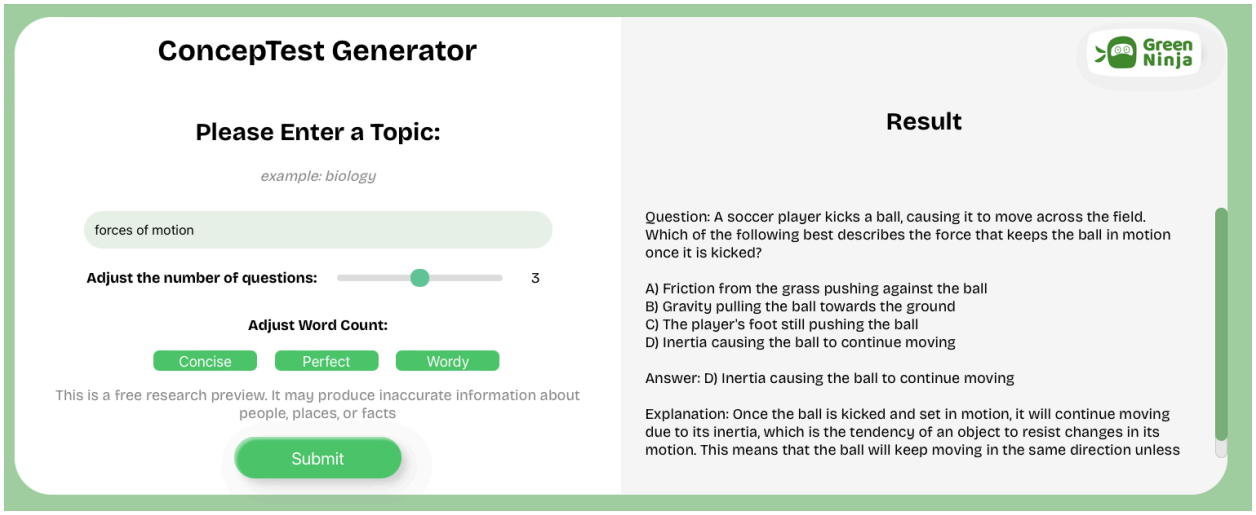


Figure 4. An existing GenAI app from Green Ninja that produces Concept-Test-styled multiple choice questions for NGSS assessments.

Recognizing that many educators have reservations about using AI (Alasadi & Baiz, 2023), we are committed to developing the GN Co-Pilot responsibly and thoughtfully, with a strong emphasis on ethical considerations, transparency, privacy, accuracy, and with humans in the loop. By collaborating with experts in AI ethics and implementing robust safety measures, we aim to create a system and platform that not only supports educational goals but also aligns

with the values and concerns of educators. Ensuring the accuracy and reliability of the GenAI systems is paramount (Yu & Guo, 2023), and we will employ rigorous techniques (prompt engineering, reinforced learning with human feedback, retrieval augmented generation, and fine-tuning) and continuous updates to maintain high standards of veracity and educational guardrails (see Appendix J.6 for technical details about the workflow to implement these guardrails). Having teachers discuss the real-world benefits and pitfalls of the GN Co-Pilot is also an essential component of the professional development (PD) for treatment teachers (see Appendix J.5). We aim to build trust and confidence among educators, making the adoption of GenAI technology seamless and beneficial.

A.5 Why it matters

Middle school is a critical stage for shaping students' lasting attitudes towards science. Engaging students in science at this stage can significantly influence their interest and proficiency in STEM fields, which is essential for developing better thinkers, a benefit for both individuals and society. We assert that using environmental problems to drive science lessons is a highly effective way to promote student learning and success. However, underserved populations are less likely to attend schools with innovative and engaging curricula, resulting in lower test scores and fewer students pursuing STEM careers (e.g., Oakes et al., 2021). This project evaluates the efficacy of Green Ninja's environmentally themed PBL curriculum on student test scores, especially those in underserved communities. Although some studies have explored the outcomes of environmentally themed instructional units lasting for a few days or weeks (see reviews by Ardoin et al., 2018; Jorgenson et al., 2019; Marcinkowski & Reid, 2019), Green Ninja is, to our knowledge, the only comprehensive middle school science program that employs

local environmental and climate solutions as the core theme. Our research aims to provide strong evidence that prioritizing environmental problem-solving can enhance overall student learning.

B. Strategy to Scale

B.1 Potential to Scale

Green Ninja was built from the ground up in 2016 for the NGSS and has a unique market niche as the only commercially available PBL curriculum for middle school. The focus on solving real-world problems has led to adoption by diverse districts ranging from the most conservative counties in Texas to the most liberal counties in California. Although more than 20 school districts have formally adopted the curriculum, growth has been slower than anticipated. This section outlines three barriers the Green Ninja team has identified based on teacher feedback and interviews with districts that piloted but chose not to adopt Green Ninja. We also describe how this *Science Roots* project will address and overcome these barriers.

Teaching using PBL barrier

Teachers recognize the benefits of PBL, including stimulating creativity, enhancing communication and collaboration skills, and promoting critical thinking (Lee & Blanchard, 2019). However, they also identify several barriers to its implementation that have hindered PBL from reaching its full potential despite its proven effectiveness:

Time: Teachers worry that planning and developing PBL activities and projects require significantly more time compared with traditional teaching methods (Krajcik & Blumenfeld, 2014; Lee & Blanchard, 2019). Implementing PBL in the classroom and assessing student projects can be time-consuming, especially when managing multiple projects simultaneously.

Assessment: Assessing student projects and providing meaningful feedback can be challenging, especially with diverse outcomes, and teachers fear that PBL might not adequately prepare

students for standardized tests. Ensuring that students can take the concepts learned from their projects and apply them to different contexts is perceived by teachers as a significant challenge. **Preparation**: Many teachers feel unprepared to implement PBL effectively due to insufficient training in PBL methodologies and strategies (Farrow et al., 2022; Pan et al., 2022). The lack of robust support systems to assist teachers in integrating PBL into their curriculum is a significant barrier. Teachers need additional guidance and resources to confidently adopt new teaching methods, especially those that involve a shift from teacher-driven to student-driven learning.

To address these barriers, we will a) provide targeted PD about how PBL works in Green Ninja with ongoing face-to-face, synchronous, and asynchronous virtual components (see Appendix J.5 for PD framework), and b) develop time-saving features using the GN Co-Pilot. The GN Co-Pilot will help with:

1. **Student feedback during project planning and implementation**. Thought starters and teacher moderated feedback to students on their projects
2. **Customized formative assessments**. Questions that prompt students to link their specific project to target science concepts so they can generalize their learning – with instant feedback and follow-up
3. **Real-time PBL teacher support through an AI assistant**. Project ideas, management strategies, and tips to teachers navigate and implement project-based learning effectively

As an example, Table 2's descriptive vignette offers a hypothetical example of ways that the training and GN Co-Pilot could help a teacher overcome barriers to teaching with PBL.

Table 2. Descriptive vignette of how teacher training and the GN Co-Pilot make PBL teaching easier.

Vignette: Meet Sarah, a middle school science teacher new to PBL. At first, Sarah was apprehensive about the time commitment and challenge of managing diverse student projects. However, after attending a summer PD session on PBL and learning how to integrate the GN Co-Pilot, she felt more confident. Sarah's top three time-saving features in using the GN Co-Pilot were: 1. Student feedback: "Helping students stay

on track with their projects is easy and fast. After students submit their project updates, the Co-Pilot provides immediate feedback, and I review the feedback, edit it if needed and send it back to students. This is great because it provides really good suggestions in near real-time." 2. Performance Assessments: "The Co-Pilot created practice assessments that linked individual projects to the science learning of the unit. This helped students tie in their learning better and prepared them for the unit assessment." 3. Support and Coaching: "Although the in-person and online training did a good job prepping me for teaching with GN, I find myself using the Co-Pilot to answer questions or help me plan. This support gives me confidence that I'm on the right track."

Market Trust Barrier

The educational publishing market is dominated by a few large legacy publishers, making it challenging for new entrants like Green Ninja to establish trust and legitimacy.

Competition with Established Publishers: To develop additional credibility with school districts, we will secure the NGSS Design Badge for some of our units. This badge, awarded by NextGenScience, certifies that our units align with the vision of the NRC Framework as evaluated using the EQuIP Rubric for Science (Next Generation Science Standards, n.d.).

Evidence-Based Documentation: To build further trust, we will demonstrate the efficacy of our curriculum through rigorous data collection and studies including the publications completed through this project's research.

Pricing Barrier

The cost of the Green Ninja curriculum has been a barrier to scaling, especially compared to competitors who benefit from economies of scale.

Higher Costs Compared to Competitors: To address costs, we will conduct a comprehensive cost analysis to develop a competitive pricing strategy. Optimal pricing models can make Green Ninja more accessible to a broader range of schools.

Cost-Reduction through GN Co-Pilot: Leveraging GN Co-Pilot can reduce development and delivery costs. Specifically, AI products such as a test bank creator, leveled readers, and teacher and student assistants will help lower costs. By automating resource development and enhancing

instructional support, GN Co-Pilot will enable us to offer more affordable solutions without compromising quality.

By addressing these barriers with innovative solutions like GN Co-Pilot and targeted teaching strategies via the PD, the *Science Roots* project has the potential to scale PBL effectively across diverse educational settings. Our approach aims to make PBL easier to teach and manage, ensuring that all students, especially those in underserved communities, benefit from high-quality, engaging science education.

B.2 Management Plan & Timeline

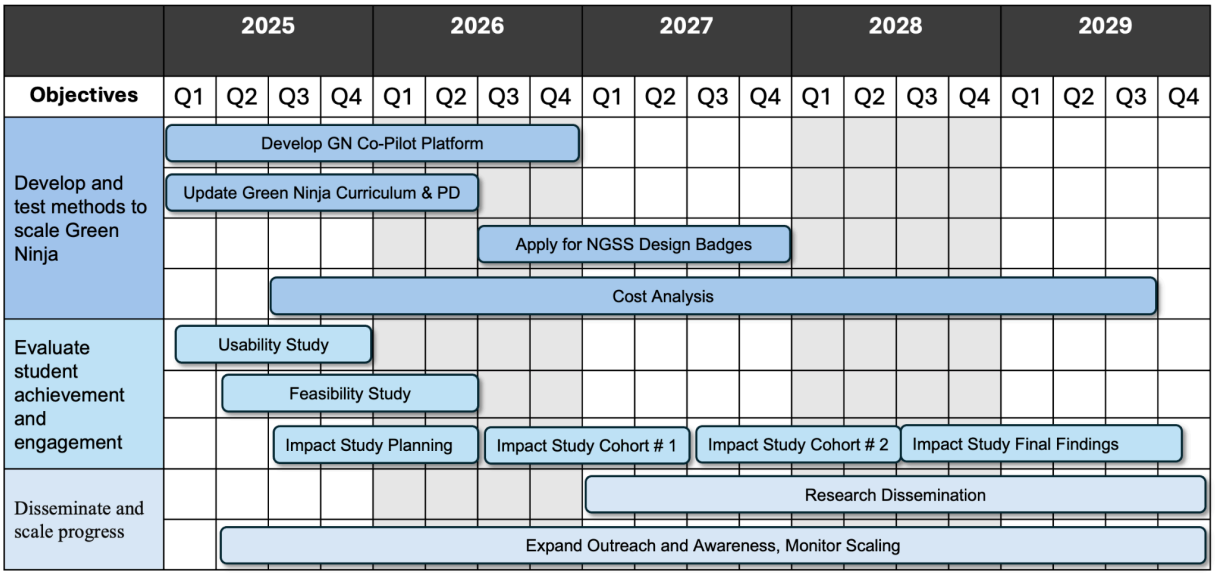
The *Science Roots* project team consists of education academics, practitioners, and technology professionals with decades of classroom experience, expertise in developing effective teacher training methodologies, and proficiency in creating and evaluating educational materials. **Roles and Responsibilities: SJSU** will be the central hub for the *Science Roots* project with [redacted] (PI) overseeing the strategic direction and [redacted] playing a leading role in updating professional training with a focus on PBL and using the GN Co-Pilot. A part-time project coordinator will manage day-to-day operations, including school recruitment, scheduling, coordination with partners, and tracking progress. **WestEd** will be the evaluation partner, conducting usability/feasibility research in the first two years and leading the main Randomized Control Trial (RCT) in Years 3 and 4, followed by exploratory studies in Year 5. During the first two years, **Green Ninja** will prioritize curriculum enhancements using GenAI methods, with continued development in Years 3 and 4 to support the impact study, and ongoing scale and dissemination work throughout the project's entirety. [redacted] from **CSUN** will serve as the curriculum lead working closely with the GenAI development team to ensure educational fidelity and accuracy of the developed product. An external **Advisory Board** will provide strategic

guidance, oversight, and support (e.g., school recruitment for impact study) through expertise in PBL, equity in education, classroom management, and GenAI. The advisory board will formally meet twice a year to review progress, and meet individually as needed with members of the project team. Appendix J.7 includes a list of advisory board members.

Meetings: To facilitate strong collaboration and establish interpersonal connections, the entire project team, including the advisory board, will meet in person once per year. These in-person meetings will include the advisory board for one day, followed by a second day dedicated to the project team. Regular bi-monthly online meetings will be conducted with the project team to discuss project objectives and ensure the strategies and activities are being met.

Project Timeline: Table 3 shows the project timeline with the major categories of activities. Appendix J.2 has a more detailed outline with start/end dates and the responsible organization.

Table 3. High-level project timeline organized by major objective.



Monitoring and Reporting: SJSU will establish a dashboard to track ongoing progress. Monthly reports on key project metrics such as task completion rates, budget adherence, and milestone achievements will be provided before monthly meetings with all partners.

B.3 Capacity of the Team

The *Science Roots* project team comprises highly qualified and experienced professionals from SJSU, CSUN, WestEd, and Green Ninja (Table 4). Each member brings expertise and a proven track record in their respective fields.

Table 4. Roles, responsibilities, and experience of key personnel.

Staff and Role	Primary Responsibilities and Relevant Experience
San Jose State University - Key Personnel	
[redacted] **Project (PI)**	[redacted] is a professor of Meteorology and Climate Science at SJSU. In 2005, [redacted] received an NSF CAREER award that helped launch research and education projects, including Green Ninja. He has developed an active education research program funded by NSF and NASA and published papers on connections between education and carbon emissions. [redacted] will lead the project ensuring strategic alignment and successful execution.
[redacted] **(Co-PI)**	[redacted] is a Professor of Geology and Science Education at SJSU and teaches Earth system science and science teaching methods. [redacted] is currently Director of the Bay Area Environmental STEM Institute (BAESI), a PD program founded in 1990 that has provided continuing education to over 2,500 teachers. She will bring her extensive experience in science education and PD to the project.
California State University, Northridge - Key Personnel	
[redacted] **(Consultant)**	[redacted] is a professor of Geoscience Education at CSUN. As a leader guiding California's NGSS implementation, he was chief writer for the 2016 *Framework for Science in K-12 Public Schools* adopted by the State Board of Education. He has provided professional training on NGSS to over 2,000 teachers. [redacted] will work on curriculum enhancement in collaboration with Green Ninja.
Green Ninja - Key Personnel	
[redacted] **Technology Leader**	[redacted] is the co-founder CTO for Engineering at Green Ninja and a technology leader with over 20 years of experience in technology development at companies including Adobe, eBay, and start-ups that use generative AI technologies. He served in various Founding, Chief, and Principal Scientist roles. [redacted] will lead strategy, execution, development and implementation of innovation and AI systems for the project.
[redacted] **Director and Scaling Lead**	[redacted] is a co-founder and CFO at Green Ninja. With two decades of entrepreneurial training and development experience, [redacted] will ensure that scale-up opportunities are maximized, focusing on the sustainable growth and impact of the project.
WestEd - Key Personnel	
[redacted]	[redacted] is a Senior Director of STEM Research and Entrepreneurship at WestEd. He has been the PI of numerous initiatives, including the NSF Center for Assessment and Evaluation of Student Learning (CAESL; $12.2M) and the i3 validation study of Making Sense of Science and Literacy ($11M). [redacted] will oversee the research and evaluation components of the project.

[redacted]	[redacted] is a Senior Research Associate in WestEd's Science & Engineering group. He is the PI of eight IES SBIR Phase I and Phase II studies and has directed the update of the NAEP Mathematics, Reading, and Science Assessment Frameworks. [redacted] will contribute his expertise in educational technology and assessment to the project.
[redacted]	[redacted] is a Research Director at WestEd, leading quantitative evaluations of educational innovations. He has over 115 publications in peer-reviewed journals and serves as an associate editor for four journals. [redacted] will lead the quantitative evaluation efforts for the project. He is currently leading an EIR Expansion evaluation.

Previous Collaborations and Successes

The *Science Roots* project is built on strong cross-institutional collaborations. The Green Ninja program originally grew out of research on student motivation at SJSU, and the two have partnered since 2010. Additionally, SJSU and WestEd have collaborated on numerous projects, demonstrating effective synergy in educational research and practice. SJSU, CSUN, Green Ninja, and WestEd have jointly submitted NSF ITEST proposals in 2021 and 2022, advancing educational innovation. [redacted] and [redacted] have co-authored conference papers and presentations, enhancing science education and PBL. All partner members have collaborated with Dr. Cordero's research team for at least 5 years and some as long as 15. The team from SJSU and CSUN has worked together to support teachers in implementing the NGSS through presentations, workshops, and publications, emphasizing inquiry-based instruction and serving high-need students. These established relationships provide a solid foundation for the project's success, and we are very excited to work together again on this project.

Financial and Institutional Resources

Each organization involved in this project is well established and committed to its success by offering access to various resources, including infrastructure and expertise. Green Ninja, which originated from academic research at SJSU, has been in the marketplace since 2018 and continues to see expansion and impact in the education sector. Green Ninja's financial stability is bolstered by a strategic partnership with World Centric Inc., a $100M annual revenue

market leader in zero-waste packaging, which provided initial support and collaborates on environmental sustainability and environmental education initiatives.

B.4 Dissemination

The Green Ninja company started as a vehicle to scale up academic research on student engagement, leveraging resources beyond those available to universities alone. However, breaking the mold of large, established publishers requires building trust and credibility. Our main strategy to build trust is through the dissemination of thought leadership – providing insightful and creative information to a wide range of audiences about our areas of expertise – PBL, equity in science education, and AI in education.

Online Tools and Resources. We will host a public version of the GN Co-Pilot for anyone to use in their PBL classroom. Alongside the GenAI, our dedicated project website will feature links to our instructional resources, digital teacher trainings, and research reports – enhanced by the creative Green Ninja art team (see Appendix J. 1).

Conferences for Educators. Recognizing the value of face-to-face interactions, we will present our findings at teacher-oriented conferences such as the National Science Teaching Association and the National Science Education Leadership Association. To reach district-level decision-makers, we will also participate in conferences such as the Council of Chief State School Officers and the American Association of School Administrators (AASA). Additionally, we will present our findings at other key events, including the Future of Education Technology Conference and the North American Association for Environmental Education conference. These events attract educators and district leaders committed to integrating innovative practices into their teaching, making them ideal audiences for our findings.

Social Media. Green Ninja has an active social media presence with over 10,000 subscribers to their main YouTube channel and 14,000 newsletter subscribers. We will share practical tips and link to our broader findings on these platforms. *Ask Matt*, an existing podcast hosted by Green Ninja about science education and the environment with over 9,000 downloads, will also chronicle our research progress and offer practical advice for educators interested in implementing PBL and environmental education in their classrooms. These venues will help us reach a diverse audience, including educators, researchers, and the general public.

Researchers. We will present our research findings at leading academic conferences including those organized by the American Education Research Association. We will publish in peer-reviewed journals like the *American Educational Research Journal* and the *Journal of Research in Science Teaching*. We also plan to write a series of white papers about how to leverage AI thoughtfully into existing educational projects. These venues will allow us to share our results broadly fostering a dialogue that can enhance the field of science education.

B.5 Broader Applications

Districts and states around the US have adopted climate literacy and environmental standards, and we expect interest in such policies to increase (EdSource, 2019; NAAEE, n.d.). A global push for sustainable education practices (UN Environment Programme, 2014) is driven in part by growing evidence that education can lead to behaviors that significantly reduce carbon footprints ([redacted] et al., 2020). In the US, teachers struggle with prioritizing implementation of these standards amongst a host of other priorities – most notably pressure to increase test scores. Underserved populations with struggling test scores feel that pressure the strongest. This project will show how environmentally-themed PBL like Green Ninja can serve both purposes effectively. Armed with the evidence of effectiveness from our research study, schools can use

the GN Co-Pilot and training resources to develop their own environmentally-themed PBL or simply adopt Green Ninja's comprehensive curriculum.

Education is also seeing a boom in AI technology. Green Ninja's focus is on developing a methodology for safe and secure AI. Our process paper will describe how we implement effective guard rails using existing technology and will inform a wide range of educational users.

C. Project Design

C.1 Conceptual Framework

The Situated Expectancy-Value Theory that underlies Green Ninja's curriculum (see Section A.3 and Figure 3) asserts that the PBL unit challenges promote a purpose for learning and build students' agency as problem solvers. These essential ingredients amplify student engagement, which influences learning. As such, it's critical that the projects in each unit are implemented with fidelity. To achieve our desired student outcomes (engagement, agency, and learning), we need to focus on supporting teachers. This support includes PD to give teachers the confidence to facilitate projects and a GenAI assistant to help augment the teacher's facilitation with differentiated and targeted support and formative assessment. The facilitation focuses on scaffolding connections between the concrete projects and the abstract science concepts (IES, 2007b), a process that works best when individualized to each project using the GN Co-Pilot.

C.2 Goals, Objectives, and Outcomes

Table 5. Project objectives, strategies, outcomes, and measures.

Strategies	Outcomes	Measures
Objective 1: Develop and Test Methods to Scale Green Ninja		
1.1. Develop, integrate and test GN Co-Pilot	- GN Co-Pilot includes AI development with guardrails - GN Co-Pilot has successful testing with students, teachers, and district staff	- Usage and feedback logs for AI system, time to value, and veracity - Level of AI integration into Green Ninja - Students, teachers, and district surveys

1.2. Improve teacher PBL confidence using GN Co-Pilot	- GN Co-Pilot is integrated into Green Ninja curriculum - Teachers feel more comfortable and confident using AI systems to teach PBL with Green Ninja	- Surveys of teacher satisfaction and confidence with PBL - Usage and feedback logs for AI system, time to value, and veracity
1.3. Overcome costs associated with Green Ninja	- WestEd completes cost analysis of Green Ninja - Competitive pricing plan is developed that includes GenAI tools	- Cost-effectiveness ratio with comparisons for other publishers - Cost plan created by WestEd
1.4. Build trust by obtaining EQuIP Design Badges and success stories	- Market trust and recognition of Green Ninja is increased - Evidence-based success and alignment with science standards is demonstrated	- # of EQuIP Design Badges - Qualitative focus groups - # of schools adopting Green Ninja
Objective 2: Evaluate Student Achievement and Engagement		
2.1 Comprehensive professional development	- Teachers are trained to effectively implement PBL including the use of AI tools - Teacher confidence and capability in using PBL and AI tools are increased	- % of teachers trained - Surveys of teacher confidence using PBL and AI tools - AI usage and feedback logs
2.2. Measure and assess the impact of Green Ninja on student achievement and engagement	- Accurate assessment of the program's effectiveness shows student outcomes are improving - Data-driven insights demonstrate student engagement and achievement	- State science test score - Analysis reports detailing the impact on student achievement and engagement
Objective 3: Disseminate and Scale Progress		
3.1 Disseminate research findings	- Visibility and credibility of the Green Ninja curriculum are enhanced among researchers and educators including district admins -Adoption of Green Ninja-informed practices in educational settings and policy frameworks increases	- Number of published articles and reports - Number of in-person and online presentations and workshops - Engagement metrics for Green Ninja-related research
3.2. Expand outreach and monitor the scaling of Green Ninja	- The number of school district adoptions increases - The number of underserved students using Green Ninja increases	- % of schools who continue to use GN after the study period - Data from outreach and engagement campaigns

C.3 Impact on Target Population

Learning with resources designed to help elevate all students. Our target population is low-income, diverse learners who are performing below proficient on the state science test. Green Ninja employs universal design for learning principles that include a wide range of learning and assessment modalities, including the PBL aspect of the curriculum. Our GN

Co-Pilot will further differentiate instruction for all students. Our initial data (discussed in Section A.3), show that Green Ninja's curriculum outperforms competitors on standardized test score growth for all students, including low-income and language learners. School districts with large populations of underserved students (e.g., Sausalito Marin City and Ravenswood City School Districts), have already chosen to adopt Green Ninja's PBL curriculum because of its engaging approach. Consequently, more schools implementing Green Ninja will benefit a greater number of students in our target population.

Teaching about topics relevant to local communities. Underserved communities in California are more likely to live in the shadow of pollution and environmental hazards and more likely to be concerned about climate change than majority white populations, yet they are also less likely to study environmental issues in their classrooms (California Environmental Protection Agency, 2021; Leiserowitz & Akerlof, 2010; Plutzer et al., 2016). Green Ninja's focus on environmental education will bring these topics to classrooms in 80 schools that are struggling for resources.

Supporting teachers with climate content. Teachers report that a lack of confidence about climate change prevents them from teaching it. Underserved districts are less likely to have resources available to provide PD to reduce that knowledge/confidence gap, so the *Science Roots* project will provide professional training for 80 participating teachers.

D. Project Evaluation

WestEd will conduct an independent evaluation of the implementation and impact of Green Ninja on 8th-grade science teacher and student outcomes (Table 6). The impact study is designed to meet WWC standards without reservation. Following extensive formative evaluation of GN Co-Pilot, the impact study will use research questions (RQs) 1 and 2 to study the impact of the intervention on student outcomes, including a WWC-acceptable state standardized test for

student science achievement. RQs 3-5 address implementation and are designed to both provide performance feedback during initial stages and document key factors that should be considered during replication and further scaling of Green Ninja. RQs 6 and 7 explore mediating and moderating effects, which will unpack how key project components and contextual factors can influence outcomes. RQs 8 and 9 are exploratory but will provide information about GN Co-Pilot implementation and the opportunity for replication and expansion based on a cost-effectiveness study of Green Ninja.

Table 6. Evaluation Research Questions, Data Sources, and Performance Measures.

Research Question	Data Source	Performance Measures
Impact analyses		
1. What is the impact of Green Ninja on students' science achievement and course performance?	State standardized science scores; Science course grades (accepted by WWC in Version 5)	Students in the treatment condition will increase their science achievement score on the state standardized assessment by 10% and increase their course grade by 0.5 points on a 5-point scale (A-F)
2. What is the impact of Green Ninja on students' perceptions of the purpose of science, empowerment as environmental problem-solvers, engagement, and interest in science, and teachers' perceptions of self-efficacy, PBL, and instructional activities?	Study-administered student survey using items from [redacted] for Purpose; the Student Agency for Learning survey ([redacted]) for Empowerment; the Activation Engagement Survey ([redacted]); and [redacted] for Interest. Teacher administered survey using the Self-Efficacy Scale, survey of PBL based on [redacted], and instructional activities (see Appendix J.3).	Students in the treatment condition will increase their purpose, empowerment, engagement, and interest in science by 15% on the student survey, and 90% of teachers will complete the PBL survey
Implementation analyses		
3. To what extent are the key elements of Green Ninja (e.g., PBL, integration across science disciplines, phenomenon-driven NGSS science) implemented with fidelity?	Green Ninja teacher and student data; and teacher biweekly survey	Teachers using Green Ninja will report using at least 70% of Green Ninja key components

4. How does implementation of Green Ninja differ across school contexts and teacher and classroom characteristics? What factors hinder or facilitate the implementation of Green Ninja?	Teacher biweekly survey; school, teacher, and student background and demographic data	100% of districts will supply extant data for all teachers and students in the study
5. To what extent does professional development and the use of the GN Co-Pilot impact teacher confidence, instructional activities, and teacher self-efficacy.	Qualitative interviews; teacher logs; Teacher Sense of Efficacy Scale (TSES)	Teachers in the treatment condition will report more confidence teaching science and increase self-efficacy by 10%
Main mediating and moderating analyses (Exploratory)		
6. To what extent does the impact of Green Ninja differ across school contexts, teacher characteristics, and student characteristics, particularly high need student characteristics?	All data for RQs 1-2, including relevant background and demographic data.	100% of districts will supply science achievement test scores and extant data for all teachers and students in the study
7. To what extent does students' perceptions of the purpose of science, empowerment as environmental problem-solvers, engagement, and interest in science and teachers' perceptions of PBL mediate their science learning?	All data for RQs 1-2	90% of students will complete the perceptions of science survey battery
Exploratory analyses (other)		
8. To what extent does GN Co-Pilot impact students' projects and their ability to connect their project to science?	Green Ninja usage studies; student surveys; teacher surveys; and qualitative interviews	90% of Green Ninja users will implement GN Co-Pilot 75% of the time
9. What is the cost and cost effectiveness of Green Ninja?	Program cost records	100% of relevant costs will be identified and included

These research questions align with the project's objectives and strategies and will be addressed using data collected from 80 middle schools in diverse settings in California.

Impact Study

We will recruit two cohorts of 40 middle schools, one starting in the 2026-2027 school year and the other starting in the 2027-2028 school year. To recruit schools into the study, we will offer free access to the Green Ninja curriculum, two days of professional training, ongoing

support during the school year, and a stipend of $1,500 for treatment teachers and $750 for control group teachers. All teachers will eventually get access to the curriculum either during the intervention year (treatment) or the subsequent year (control).

Table 7. Demographics in counties where County Offices of Education have pledged to help with recruitment.

County	Number of Middle Schools	All Middle School Students		Title 1, Low SES students	
		Num Students	% Below Standard	% of students	% Below standard
Fresno	538	14,570	22.4	76	26.6
Los Angeles	3,173	95,538	18.8	71	22.8
San Joaquin	571	11,109	20.1	63	24.6
San Mateo	247	6,044	13.8	36	25.6
Santa Clara	504	17,096	13.1	39	23.7

We will randomly assign each school in each cohort to the treatment (Green Ninja) or control conditions, blocking on prior year school-level % of students below the standard in science, Title I status, and % of White students in the school to ensure baseline equivalence at the unit of randomization. Overall, half (n = 40) of the 80 middle schools in the study will be randomly assigned to have their 8th-grade science teachers participate in Green Ninja implementation, and the other 40 schools will continue with business-as-usual 8th-grade science curriculum and instruction. Recruitment will target schools with a high percentage of low-income students but will ensure an appropriate range of demographics for generalization. For a priori study design purposes, we assume one 8th-grade science teacher per school, though some may have more. Thus, this design includes approximately 80 teachers (40 in each condition) and 8,000 students (100 per teacher) for the impact study. [redacted]

WWC Acceptable Outcome: Student Science Achievement. To measure students' outcomes for RQ 1, WestEd will use the California 8th Grade End-of-Grade Science Assessment, which is considered valid and reliable by the WWC 5.0 standards. The assessments were developed by a rigorous standards setting process to ensure high validity and reliability.

Students' course grades provide a supplemental, exploratory measure of student science achievement. Because grading practices differ from teacher to teacher, we do not interpret course grades as a precise measure of student learning. Rather, we consider them a marker of course performance, which is policy relevant and a strong indicator of future academic success ([redacted]). The WWC Study Review Guide Version 5.0 includes course grades as an eligible outcome for review as long as the same scale of GPA is used across sites.

Outcome: Students' perceptions of the Purpose of Science, Empowerment as Environmental Problem-Solvers, Engagement, and Interest in Science. WestEd will develop a student survey that combines items and domains from psychometrically valid surveys of the core perception constructs (see Appendix J.3 for example surveys and items). We will create, pilot test, and refine the survey through cognitive interviews with 8th-grade students and reliability and validity analyses prior to the start of the RCT. Domain specific scores (Purpose, Empowerment, Engagement) will be modeled separately in analyses for RQs 2, 6, and 7.

Outcome: Teacher Self-Efficacy. [redacted]

Outcome: Teachers' Perceptions of PBL. WestEd will administer a survey based on [redacted] The PBL survey includes eight items that measure teachers' perceptions about guiding

small groups of students to develop solutions to a given problem, encouraging collaborative learning, and strengthening critical thinking and communication skills (α = 0.96).

Outcome: Teacher Instructional Activities. To measure the quality of instructional activities (RQ 5), WestEd will administer a teacher log three times across the year during the impact study. Prior studies of teacher logs indicate that they can be a valid and reliable measure of instruction ([redacted]). The log will include the following measures adapted from a RAND study of inquiry-based instruction ([redacted]): inquiry-based practices that actively engage students and promote problem-solving skills (α = 0.83), inquiry-based activities that facilitate critical thinking (α = 0.77), discussion (α = 0.74), and developing conceptual understanding (α = 0.58). Together, these measures capture the types of instructional activities expected in Green Ninja classrooms (see Appendix J.3 for the complete set of items). WestEd will also perform at least 40 observations of treatment and control classrooms (~50% of the sample) during the impact study to observe implementation and differentiate between treatment and control conditions.

D.1 Meeting WWC Standards Without Reservations

We will evaluate the impact of Green Ninja based on a school-level randomized controlled trial designed to meet WWC 5.0 standards without reservations ([redacted]). We will assign participating schools to the treatment or control condition using blocked random assignment. Blocks will consist of school-level demographic information and prior school-level science assessment outcomes (see Appendix J.3).

After randomization, Green Ninja will enroll the treatment group into its PD sequence. All teachers in the treatment schools who plan to teach 8th-grade science in the coming year will

be included. Treatment schools will receive Green Ninja's PD program, and those teachers will have full access to all of its resources, including the GN Co-Pilot.

The control schools will administer business-as-usual science curriculum and will not have training with or access to Green Ninja. Their PD will include regular offerings from their district, county, or state. We will use school-level random assignment since schools typically adopt Green Ninja as a school-wide curriculum. Further, a teacher-level assignment would raise the threat of contamination, as teachers in a school may discuss, view, and share instructional materials and strategies. Also, based on prior school-level randomized studies ([redacted]), including current EIR studies by the evaluation team ([redacted]), we expect minimal school-level attrition during the impact study. See Appendix J.3 for attrition scenarios and impact on power.

The analysis of the intervention's impacts will use an intent-to-treat (ITT) approach—schools and their teachers and students will be retained in their originally assigned groups. We will collect student rosters at the start of each school year to identify students in the ITT student impact sample. To get a comprehensive assessment of the instructional activities students experience during the study, the primary impact analysis for teacher instructional activities will estimate ITT effects. Given that the proposed evaluation is based on a school-level RCT that is expected to have low cluster-level attrition and a student analytic sample where joiner bias is not a threat, the evaluation has the potential to produce strong evidence about the impact of Green Ninja. Students who join the school after randomization will not be included in the analytic sample. In addition, based on research ([redacted]) and WestEd's prior RCT experience, we anticipate manageable levels of student attrition during the study (i.e., less than 20%) and minimal differential student attrition across conditions (i.e., less

than 5 percentage points), so the student impact analyses as designed will meet WWC standards without reservations ([redacted]).

D.1.a Generalizability and Scalability. In partnership with SJSU, WestEd will evaluate Green Ninja across a large number of economically disadvantaged schools, particularly in high-minority settings. Findings based on this diverse study sample will provide valuable guidance for future replications of program implementation as the program scales to similar underserved populations across the country. To inform generalizability, we will conduct moderator analyses (RQ 6) to assess the extent to which characteristics of students, teachers/classrooms, and schools impact the outcomes. Results from these analyses will identify where the program is most effective and how it can be improved to be effective across all characteristics. Appendix J.3 lists potential moderators and a corresponding analysis plan. We will explore how implementation and instructional practices, student perceptions, engagement, and empowerment mediate the direct effects on students' science achievement.

D.1.b Cost Effectiveness. The evaluation includes a cost analysis based on the Resource Cost Model ([redacted]) to provide information about the cost of implementing Green Ninja, including associated PD and support, and whether it is cost effective relative to the Business-as-Usual (BAU) condition. We will use the "ingredients method" to identify Implementation and PD costs in both the Green Ninja and BAU conditions ([redacted]). We will identify the costs associated with each component of the program, distinguish start-up costs from ongoing costs, and convert total costs to per-student costs. We will then combine the cost information and effect size estimates to describe the impact of Green Ninja on a per dollar basis following the most up-to-date recommendations ([redacted]).

D.2 Strategies for Replication

Our research questions explore how, when, and for whom Green Ninja works. Rigorous moderation and mediation analyses will help Green Ninja identify barriers that hinder scaling so the team can address them. There is then ample opportunity to test out these revisions using the materials and methods from this study. Green Ninja targets district-wide adoptions, so the program can scale very quickly and provide new opportunities for future replication studies. Green Ninja will prioritize continued research and improvements using its own funds and potentially through other partners and grant monies.

D.3 Components, Mediators, Outcomes, and Acceptable Thresholds of Implementation

The proposed evaluation design is informed by clearly articulated key components, mediators, and outcomes of Green Ninja as depicted in the conceptual framework (Figure 3) and logic model (Appendix G). The impact analyses (RQs 1 and 2) will be based on valid and reliable measures described above and in Appendix J.3. The evaluation will include moderator analyses (RQ 6) and mediator analyses (RQ 7) to explore the relationships among implementation context, intermediate outcomes, and student achievement outcomes. Implementation context data will be collected from multiple sources, including artifacts from teacher trainings (e.g., sign-in sheets and agendas) to determine participation and coverage, observations, qualitative interviews, and monthly teacher logs describing teaching activities using Green Ninja and other curricular materials. Acceptable thresholds of implementation will be defined for PD measures, teacher logs, and observations during the formative evaluation stages (see D.4). Thresholds will likely be based on the percentage of weeks in which teachers attended and completed the expected number of activities prescribed in the administration model.

As we monitor these thresholds, we can determine if the program is implemented with fidelity and what potential effect the level of fidelity may have on outcomes.

D.4 Formative Evaluation, Performance Feedback, and Periodic Assessment of Progress

The first 1.5 years of the evaluation will provide formative performance feedback and periodic assessment of progress. Usability and feasibility studies with GN Co-Pilot will build towards the impact study, and be guided by corresponding research questions (see Appendix J.4).

Table 8. Summary of the samples and timeline for each major evaluation component.

	Usability	Feasibility	Impact
Timeline	Summer 2025	2025-2026 school year	2026-2028
Teachers	10	10	80
Students	10	1000	8,000

Usability Study

In summer 2025, WestEd will conduct multiple rounds of usability research to iteratively test new program components, features, and content that were built in the winter/spring by SJSU and Green Ninja, particularly GN Co-Pilot. For each round, at least 5 teachers and 5 students will participate with a minimum of 10 teacher and 10 student participants total. Research has shown this is a sufficient number of testers to identify major issues, and that testing with additional users provides diminishing returns ([redacted]). The sample will be intentionally balanced for demographics and contexts. WestEd researchers will guide participants through the relevant tasks and ask them to "think aloud" and explain their thought processes as they go. At the end of each session, WestEd will interview participants about their overall experiences, the ease of use of the activities, and their understanding of the tools and content presented. After each round of usability, WestEd will summarize findings with actionable recommendations to SJSU and Green Ninja.

Feasibility Study

Starting in fall 2025, WestEd will conduct a feasibility study to evaluate whether teachers can use the PD components, GN Co-Pilot, and materials, and translate them into practice in an authentic education setting. We will also explore how teachers envision ideal usage of Green Ninja to improve PD and administration models for the Implementation Study. WestEd will complete the feasibility study with 10 teachers, each with one full class of students. Teachers will implement Green Ninja and use GN Co-Pilot in their classrooms. We will focus on measuring the veracity, time-to-value, and usability of the new GN Co-Pilot AI system. WestEd will capture data about implementation using interviews and observations. We will collect feedback from teachers through a bi-weekly log and a final interview. Qualitative and quantitative reports of findings will provide actionable recommendations to SJSU and Green Ninja throughout the year. Additionally, WestEd will work with teachers to have their students complete the Purpose, Empowerment, Engagement, and Interest measures for psychometric evaluation.

Performance Feedback and Periodic Assessment of Progress. During this formative period, the team will collaboratively develop the ideal administration model that can work best for teachers in the study. WestEd will also create measurable thresholds for acceptable implementation based on teacher feedback and findings from the implementation study.

The formative evaluation will also provide structure for consistent, periodic feedback to SJSU. All parties will participate in virtual, bi-monthly meetings to discuss upcoming project goals and review progress to date on development, testing, and other project components, as detailed in the Project Timeline (Table 3) and Management Plan by Project Objectives, Outcomes, and Performance Measures (Table 5). WestEd will be responsible for monitoring and tracking all research activities and reporting to SJSU to make sure all goals stay within their scope and timeline.