Check for updates

Science Education

WILEY

Science Education

RESEARCH ARTICLE OPEN ACCESS

Impact of Teachers With Research Experiences: Student Gains in STEM Career Awareness, Perception of Value of STEM Learning, and Persistence in STEM Course Tasks

John Keller[1,2] | Sanlyn Buxner[3] | Dermot Donnelly-Hermosillo[4] | Elsa Bailey[5] | Martyna Citkowicz[6] | Larry Horvath[7] | Dan Moreno[3] | Melissa Yisak[6] | Bo Zhu[6] | Eleanor Fulbeck[6] | Deidre Sessoms[8] | Stamatis Vokos[2] | Charlotte Chen[6] | Max Pardo[6]

[1]Astrophysical and Planetary Sciences, Boulder, Colorado, USA | [2]Center for Engineering, Science and Mathematics Education, California Polytechnic State University, San Luis Obispo, California, USA | [3]Teaching, Learning, and Sociocultural Studies, University of Arizona, Tucson, Arizona, USA | [4]Chemistry and Biochemistry Department, California State University Fresno, Fresno, California, USA | [5]Director and Principal, Elsa Bailey Consulting, Pleasant Hill, California, USA | [6]Human Services Division, American Institutes for Research, Arlington, Virginia, USA | [7]Secondary Education, San Francisco State University, San Francisco, California, USA | [8]Undergraduate Studies in Education, California State University Sacramento, Sacramento, California, USA

Correspondence: John Keller (john.m.keller@colorado.edu)

Received: 25 April 2022 | **Revised:** 15 July 2024 | **Accepted:** 14 October 2024

Funding: Prior to receiving support from the National Science Foundation, influential preliminary work on this research investigation was supported by the Noyce Foundation and 100Kin10. This material is based upon work supported by the National Science Foundation under Grant Nos. 1660839, 1660810, 1660777, 1660715, 1660658, and 2140288. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Keywords: comparative | comparison | follow-up | long term impact | postinterventional | QED | student achievement data | student survey outcomes | teacher research experience

ABSTRACT

Research Experiences for Teachers (RET) programs are a burgeoning approach to engage teachers in STEM (science, technology, engineering, mathematics) research that they can translate into their K-12 classrooms. Despite an increase in studies of RETs, there is a need for comparison of RET and non-RET teachers' student outcomes. This mixed methods, quasi-experimental comparison study, using a revised third-generation activity theory framework, investigates how an RET program for preservice and early career STEM teachers impacted participating teachers and their students up to 8 years after RET participation. Specifically, we conducted a matched comparison of student achievement data from students of nine RET teachers versus many non-RET comparison teachers within the same districts ($n = 830$–1132 students). We also investigated student and teacher perceptions of classroom practices through surveys ($n = 576$ students) and interviews (15 teacher interviews). Omnibus tests revealed no statistically significant differences by treatment in math or science achievement. However, students of the RET teachers reported stronger perceptions of STEM career awareness, greater value for learning STEM subjects, and a greater propensity to persist in STEM course tasks (three of the five constructs measured). This was consistent with teacher interview responses in which RET teachers spoke about STEM career awareness in a broader context for understanding the value of STEM in society, and also discussed struggles in research and attempts to bring this mindset to their students, which may have resulted in greater student engagement in their courses. Implications for teacher education and for supporting science and engineering practices in STEM classrooms are discussed along with recommendations for further research on the impacts of RET programs guided by a revised third-generation activity theory framework informed by this work.

1 | Introduction

Curriculum standards such as the Next Generation Science Standards (NGSS Lead States 2013) place an important focus on K-12 students enacting science and engineering practices (SEPs) such as planning and conducting investigations, analyzing and interpreting data, and using mathematics and computational thinking. Teachers, in many cases, have had limited opportunities to experience such practices outside of the classroom, yet are responsible for developing lessons and activities to support these practices (Feldman and Özalp 2019; Haag and Megowan 2015; Harris, Sithole, and Kibirige 2017; Kang et al. 2016). As a result, Research Experiences for Teachers (RET) programs in Science, Technology, Engineering, and Mathematics (STEM) have received increasing attention as a potential means to engage teachers in experiencing STEM research practices outside of the classroom while also informing teachers of entry points and careers in STEM that they can integrate into their classroom activities (Davidson, Jaber, and Southerland 2020; Shanahan and Bechtel 2020; Stieben, Pressley, and Matyas 2021; Yang, Liu, and Gardella 2020).

Broadly speaking, RET studies have investigated teachers' beliefs about: (a) themselves; (b) the nature of science and inquiry; (c) changes in teaching practice; and their knowledge about: (d) science content; and (e) science practices (Feldman and Özalp 2019: reviewed 25 empirical RET studies). These RET studies predominantly use self-report measures, which can provide valuable insight on RET participant beliefs, perspectives, and insights. However, self-report measures in isolation as a data collection method have also been critiqued for having reliability issues in accurately determining what happens in day-to-day classroom interactions and how those interactions impact student learning (Feldman and Özalp 2019). Therefore, it is desirable to have more RET studies that include measures of student learning, comparisons with teachers without an RET, and longitudinal insight on how RETs may impact teachers across the span of their careers. Three studies that include student learning outcomes find that students of RET teachers show: (a) significant pre/post science conceptual and scientific literacy learning gains (Ragusa and Juarez 2017); (b) a positive association between student understanding of interdisciplinary science concepts and length of RET placement (Yang, Liu, and Gardella 2020); and (c) superior performance of students of RET teachers compared to students of non-RET teachers in state science exams (Silverstein et al. 2009).

Ragusa and Juarez (2017) report statistically significant gains for students of RET teachers in conceptual understanding on concept inventory measures developed by the research team that were specific to grade levels for science, computer science, and engineering. There were also gains in subject-specific science literacy (chemistry literacy, physics literacy, etc.; where the research team developed and validated a qualitative reading measure). However, the number of students in the analysis is not reported. Yang, Liu and Gardella (2020) find that inquiry instruction, supported through RETs, promotes greater student understanding of interdisciplinary science concepts (20 multiple-choice items focused on NGSS Cross-cutting Concepts adapted from existing literature; $n = 5581$), but argue that their findings should be considered exploratory. Neither of these two studies include data on comparison teachers. Lastly, Silverstein et al. (2009) find that when compared with students of non-RET teachers, students of RET teachers (who participated in research for more than 2 years) pass the New York state science exams (Biology, Chemistry, and Earth Sciences) at a 10.1% higher rate than non-RET teachers' students ($p < 0.049$; $n = 43{,}310$). Based on the limited number of studies that investigate student learning related to RETs, further research is needed to understand whether there are observable differences in student outcomes for those taught by RET teachers relative to those taught by non-RET teachers.

The present study informs some of the aforementioned gaps and extends the existing research literature by providing a matched sample comparison of practicing high school STEM teachers who completed a particular RET, either as preservice teachers or as early career teachers, relative to outcomes for practicing teachers without an RET. Data sources include student Science and Math achievement data, surveys of students' perceptions of their learning, and interviews of teacher perceptions of their classroom practices. Data sources were collected in the 2018–2019 academic year, which was 3 to 8 years after RET alumni participated in the research program. The study investigates the effects of being taught by RET teachers on student academic and nonacademic outcomes and includes data from both RET and non-RET teachers and their students. We will first provide a review of prior work to study impacts of RETs.

1.1 | Impact of Teacher Research Experiences

Research Experiences for Teachers are defined as experiences designed for preservice or inservice teachers to engage in research under the guidance of a STEM researcher. A review of the literature shows that RETs should include concurrent and potentially long-term support to help preservice or inservice teachers translate their research experience into classroom practices (Blanchard and Sampson 2018; Krim et al. 2019). To be fully inclusive, RETs should also recognize and value the expertise that all stakeholders bring (mentors, teachers, research participants) in terms of science, research, and teaching (Shanahan and Bechtel 2020).

Previous research has measured the impact of RETs primarily on teacher level outcomes (Feldman and Özalp 2019) with some studies including additional focus on mentor scientists and/or students (Shanahan and Bechtel 2020; Snitynsky, Rose, and Pegg 2019; Yang, Liu, and Gardella 2020). In terms of teacher measures, studies generally find that teachers who participated in an RET report improved teaching practices, improved beliefs about themselves and knowledge about the nature of science and inquiry, and a better understanding of science career pathways (Snitynsky, Rose, and Pegg 2019; Storm and Lichtenstein 2019; Warburton et al. 2019). For example, in one RET study (Stieben, Pressley, and Matyas 2021), teachers report feeling more confident in leading their students to conduct scientific research, have increased connections with the scientific community, and are more likely to report integrating effective pedagogies such as inquiry and project-based approaches with corresponding assessment methods. In particular, the teachers report increasing their focus on students generating questions and pursuing their own investigations, which aligns with the SEPs of the NGSS (NGSS Lead States 2013).

Prior research also has reported equivocal findings regarding increases in RET participants' science content knowledge, using both quantitative and qualitative measures (Feldman, Divoll, and Rogan-Klyve 2013; Westerlund et al. 2002), with some finding no increases in conceptual understanding (Luera and Murray 2016). Studies also show measurable and significant changes in teaching practice via pre/post RET classroom observations using the Reformed Teaching Observation Protocol (RTOP) (Amolins et al. 2015; Southerland et al. 2016; Yezierski and Herrington 2011). These studies find that research experiences focused primarily on improving teacher learning, as opposed to research experiences with a central goal of adding to the body of science knowledge, were more likely to impact teachers' classroom practice (Southerland et al. 2016), and that observing measurable differences in teacher practice can take as long as 2 years, including the research experience and adapting RET materials to their classroom (Yezierski and Herrington 2011). Larger class sizes typically found in high schools challenge teachers' abilities to implement changes in classroom practice because it is more difficult to do inquiry-based instruction. Furthermore, teachers with many large classes may also know less about their students. Therefore, the transformation of increases in inquiry knowledge to inquiry instruction is not guaranteed (Yang, Liu, and Gardella 2020). These findings illustrate important factors in determining the impact of an RET on teachers' practice and their effect on student outcomes.

Despite the importance of understanding the impact of RETs, few studies have tracked student outcomes longitudinally. Most report on student outcomes via indirect accounts from teachers rather than including student measures within their research designs. Beyond the studies already discussed that directly investigate conceptual outcomes for students of RET teachers (Ragusa and Juarez 2017; Silverstein et al. 2009; Yang, Liu, and Gardella 2020), we identified only two studies that include measures for, and direct accounts from, students of RET teachers (Autenrieth, Lewis, and Butler-Purry 2018—student survey; Reynolds, Yazdani, and Manzur 2013—informal student comments). These two studies find that students of RET engineering teachers show increased subject awareness around engineering and increased interest in pursuing an engineering career. It is apparent that more studies that provide insight on the experiences of students of RET teachers would be valuable.

There is considerable variation in how studies have investigated the impact of RETs, with important differences in critical study features such as research design, sample size, and participant types. Given the small number of participants common in RETs, case studies are a typical empirical approach. As such, the generalizability of many studies is challenging given the unique features of individual RET programs. Research designs that include larger samples (e.g., survey studies) usually focus solely on self-report measures with nonvalidated items and lack insights on classroom practice (Feldman and Özalp 2019; Yang, Liu, and Gardella 2020). In addition, studies vary in terms of participants included, with some focusing solely on teachers, some on teachers and scientists, and others including teachers and their students. Overall, it is challenging to make valid, reliable, and generalizable claims about the impact of RETs on teaching effectiveness.

The Collaborative Around Research Experiences for Teachers (CARET), a group of representatives from over a dozen US research programs, convened in 2017 to discuss approaches to assess RETs and other research programs for students and teachers. After reviewing 307 papers CARET developed a research-based model for illustrating the relationship between elements in a teacher's RET communities of practice, professional learning communities, and outcomes for teacher and their students. RET's often bring teachers into the Community of Practice (CoP) of scientists, giving them direct exposure to scientists' culture, environments, attitudes, methodologies, and tools. This model highlights that impacts of RETs on teachers should also result in observable changes in their students (Krim et al. 2019). In response to the ideas presented in the CARET model, we sought to include not only teacher data but also student outcome data in our research. In addition, Krim et al. identified a need for more research of RETs to be based on explicit theoretical frameworks. Our research team identified Third Generation Activity Theory (Engeström 2015; Wade-Jaimes, Cohen, and Calandra 2019) as an appropriate theory to ground our work. During subsequent data analysis, our team further refined this framework to inform the work presented below.

1.2 | Description of RET Associated With This Study

The RET investigated in this study, the STEM Teacher and Researcher (STAR) Program, places prospective preservice and early-career STEM teachers in paid summer STEM settings for 8–10-week research experiences (Table 1). Participants in the STAR Program are hereafter referred to as *STAR Fellows* while engaged in the formal program and as *STAR alumni* after participation is completed. While not all of these STAR alumni ultimately became teachers, those who did enter the classroom are hereafter referred to as *RET teachers*.

Eligible applicants include: (a) students affiliated with one of the 22 campuses within the California State University system that organized the program; or (b) NSF Noyce Scholars from campuses across the US (Richardson 2016). Described in more detail in Supporting Information S1: Section S1, Noyce Scholars are STEM majors who receive scholarship funding in exchange for a commitment to teach in a US high need district after graduation. For the purposes of this study, schools with at least 50% of students qualifying for free-or-reduced-price lunch have been defined as high need. Approximately half of STAR alumni fit into this latter Noyce category.

STAR Fellows can start as early as the summer before completion of the undergraduate degree and can return for up to two additional summers at any time up until they complete their second year as K-12 instructors of record. STAR research placements have primarily involved partnerships with national

TABLE 1 | Key elements of the STAR Program, the RET involved in this study.

RET Components	Description
Eligibility/Recruitment	Preservice candidates who have completed at least their third year of college but have not yet become K-12 instructors of record, or in-service teachers who began the program as preservice candidates but have not yet started their third year of teaching as K-12 instructors of record.
	Affiliated with a California State University campus and/or an NSF Robert Noyce Teacher Scholarship Program.
Selection/Placement	Applicants submit applications with short answer questions about research and teaching and one recommendation form submission.
	Initial screening by STAR Program staff vetting for interest in K-12 teaching followed by matching process heavily influenced by laboratory site and mentor input.
Location	Research placement at either national laboratory (NASA, NOAA, NSF, DOE, DOD, USGS) or a university campus.
Duration	8–10 weeks during summer months.
Compensation	Research stipend paid to participants to cover moving expenses and housing and to provide summer pay.
Summer research experience components	Mentored research typically working with an individual mentor or research team, often contributing to a predefined research project.
	Commitment to work 40 h/week with 36 h/week conducting research and 4 h/week participating in STEM Education Workshop sessions.
	Participation in weekly STEM Education Workshop brings together cohorts of 4–12 other RET Fellows to provide support and explore topics around integration of research experience into classroom practice.
	Attendance at summative research conference.
Summer deliverables	Research poster or presentation given at summative research conference.
	K-12 lesson plan/activity with some connection to STAR Fellow's research experience.
	Participation in program evaluation.
Alumni support	Teacher professional development workshops offered during summers following STAR.
	Travel awards to present at professional research and/or teaching conferences.
	Informal support from STAR Program staff and workshop leaders in searching for teaching positions and during first years of teaching.

Note: Table highlighting key programmatic elements of the RET in this study.

laboratory facilities (Table 2) along with a small number of placements with university research groups and nonprofit research and development laboratories. In its 15-year existence up through and including summer 2021, the STAR Program had placed 590 individuals in 820 research placements.

STAR Fellows participate in a required weekly STEM Education workshop in small cohorts of 4–12 Fellows. The STEM Education workshops are led by K-12 master teachers or university teacher educators who have extensive experience in both research and teaching, as well as deep expertise in facilitating adult learning. The focus of the STEM Education workshop is twofold: (a) to help Fellows reflect on the everyday STEM practices of researchers and research groups and interpret those as concrete instantiations of nature of science (NOS) abstractions (e.g., Erduran 2017; Latour 1993; Lederman, Lederman, and Antink 2013); and (b) to assist Fellows as they translate the research experience into possible future classroom practice, even for participants who may have not yet started a postbaccalaureate teaching credential program (which is the primary teacher preparation route in California, the state from which most STAR Fellows are recruited). With regard to NOS, workshop discussions centered around the differences between observation and inference, the importance of human imagination and creativity in science, that science is practiced within broader cultural contexts, and that scientific knowledge is often subjective or theory-laden and never absolute (Lederman, Lederman, and Antink 2013). In addition to completion of the research experience, Fellows are required to produce two deliverables: a professional poster about their research project (appropriately vetted by their research group or lab) and a lesson (or lessons) that engages students in open-ended investigations based on their summer research experience, which is appropriate for the grade-band they will teach. In most cases, Fellows also practice teaching by facilitating peer–peer demonstration of these lessons. Finally, Fellows present their posters in a research conference that brings all Fellows to one location to celebrate a successful culmination of the program.

TABLE 2 | Lab sites participating in RET program.

Lab site characterization	Lab site
Department of Defense (DOD)	Air Force Research Laboratory
Department of Energy (DOE)	Lawrence Berkeley National Laboratory
	Lawrence Livermore National Laboratory
	National Renewable Energy Laboratory
	Oak Ridge National Laboratory
	Pacific Northwest National Laboratory
	Sandia National Laboratory
	SLAC National Accelerator Laboratory
National Aeronautics and Space Administration (NASA)	Caltech/Jet Propulsion Laboratory
	NASA Ames Research Center
	NASA Armstrong Flight Research Center
	NASA Goddard Space Flight Center
National Oceanic and Atmospheric Administration (NOAA)	NOAA Earth Systems Research Laboratory
	NOAA Southwest Fisheries Science Center
National Science Foundation (NSF)	Laser Interferometer Gravitational-Wave Observatory (LIGO)
	National Center for Atmospheric Research (NCAR)
	National Ecological Observatory Network (NEON)
	National Optical Astronomy Observatory (NOAO)
United States Geological Survey (USGS)	Channel Islands Field Station
University Campus	California Polytechnic State University San Luis Obispo
	Fresno State University
	Sacramento State University
	San Francisco State University
Other Research Institutions	Biosphere 2
	Estuary and Ocean Science Center
	Monterey Bay Aquarium Research Institute
	SETI Institute
	Space Sciences Laboratory

Note: Table of partner organizations and laboratory sites where STAR Fellows were placed in summer research experiences between 2010–2016.

STAR staff also provide professional development opportunities and support to alumni, including participation in professional conferences and summer professional development workshops on both STEM teaching and computer coding.

1.3 | Conceptual Framework

To address the call for research on RETs guided within more explicitly-stated theoretical frameworks as suggested by Krim et al. (2019), we further enhanced our theories of situated learning/communities of practice with Third Generation Activity Theory, as presented by and Wade-Jaimes, Cohen and Calandra (2019). This model was chosen to address the fact that both research experiences and teaching occur in complex settings, and we were looking for a framework to attend to complexity and interactions that occur both in and between these settings.

Classrooms are complex habitats (Doyle 2013) where teachers and students work together to navigate a myriad of decisions that influence the course of any particular lesson, class session, or school term. Teacher decision-making is based on three practicality criteria: (a) ease of implementing an action; (b) content and procedures of the change; and (c) whether the change aligns with what they believe. Teachers both consciously and subconsciously weigh the ease with which a proposed action might be implemented, along with the congruence and cost of proposed actions (Doyle and Ponder 1977; Janssen et al. 2013). In terms of ease of implementation, teachers must be explicitly taught both the content and the procedures of the intended change. Simple statements of value—saying it would be better to teach in a different way—are insufficient for teachers to implement a change, even if they agree with the value of that change. Teachers view actions as congruent when the implementation and intended outcomes of the action match the perceived needs and contextual variables of the teachers' classroom and school settings. The costs associated with a practical action speak to teachers' perceptions that the returns on their investment of time, effort, and energy will be sufficiently greater than the expenditures. Elements of the RET program investigated in this study address these three practicality criteria, and a Third Generation Activity Theory framework serves as a useful tool to design and analyze an investigation to address these criteria.

Activity occurs when a subject uses tools (e.g., experiences, curriculum) to work on an object (e.g., skills, dispositions) to achieve an outcome (Vygotsky 1978). These activities are the source of human development and learning, inextricably situated in time and place, and informed by the history and culture of the community in which the activity takes place (Leont'ev 1978). The Third Generation Activity Theory adds *community* as well as *rules* and *division of labor* to the activity system. It also allows for the contribution of multiple activity systems acting on the same object and acknowledges the existence of contradictions between constituents of each activity system as well as between systems acting on a shared object (Engeström 2016).

In the combined activity system that we developed for this study (Figure 1), a preservice or early career teacher (the subject) can use the tools provided by the RET program to enhance their own knowledge, skills, and dispositions in science and science education (the object) toward the outcome of teaching with, and engaging their future students in, authentic scientific research tasks using the SEPs of the *Framework for K-12 Science Education* (National Research Council 2012) and the *NGSS* (NGSS Lead States 2013). RET Fellows engage in authentic research tasks in laboratories and field settings. RET program mentors and workshops explicitly connect these authentic research experiences to classroom content, pedagogy, and specific teaching practices, which addresses the instrumentality criterion of teachers' decisions to use the research content and those practices in the future. By the end of an RET Fellow's experience, alumni should have the awareness of practices used by professional scientists and engineers to conduct research, a preliminary ability to use those practices to conduct research themselves, the knowledge of why those practices should be implemented in a K-12 science classroom, and an understanding of how to start implementing the practices in their own classroom, supported by other experiences from their credential program. Additionally, RET alumni may have dispositions toward the aforementioned. To develop that knowledge and those skills and dispositions, the RET was designed in such a way that the rules, community, and division of labor in the RET program and the research settings in which the RET alumni work facilitate this objective (Engeström 2016). Instructional changes that teachers have the skills to implement, and that align with teachers' dispositions toward classroom practice, are far more likely to be successfully implemented and maintained over the long term (Doyle and Ponder 1977; Janssen et al. 2013).

Within this activity theory model, teachers' knowledge, skills, and dispositions are also influenced by multiple other activity systems. Most observable in this investigation is the school activity system. Teachers' school activity systems may play a larger role than the RET activity system in determining the practicality of implementing changes to instructional practices. Tools (e.g., supplies, equipment, physical classroom structures, etc.) vary widely between schools, even within the same districts, as do teaching practices and community values. As informed by the RET activity system, alumni may have the knowledge, skills, and dispositions necessary to implement

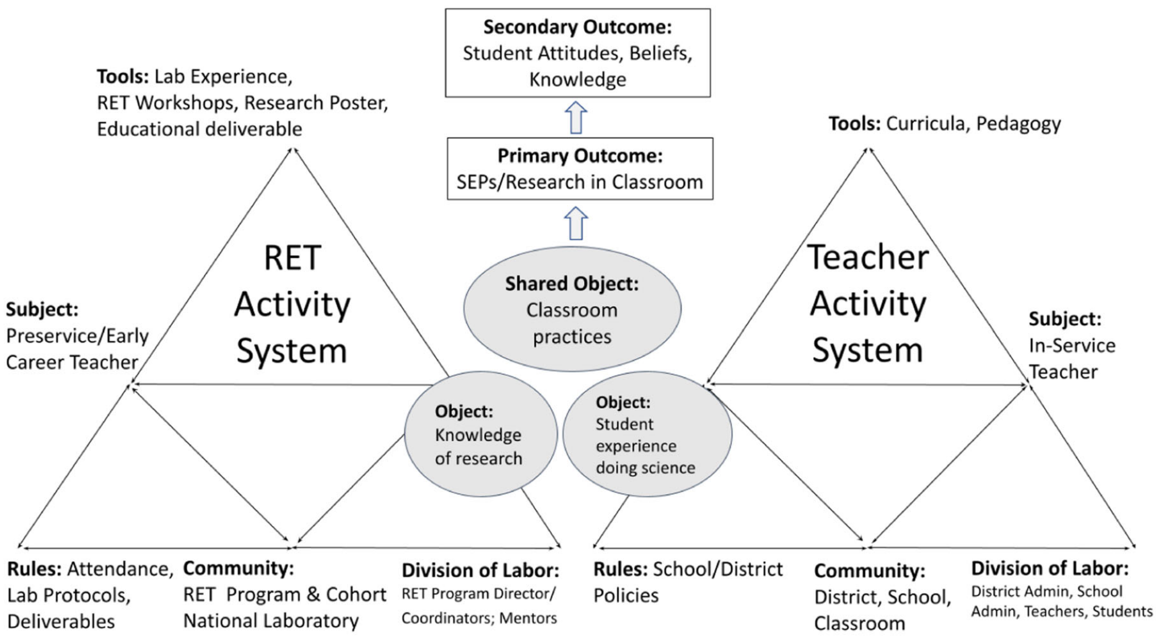

FIGURE 1 | Conceptual framework. *Source:* This figure demonstrates the conceptual framework for conducting research on RET programs developed through this study. The framework is an extension of Third Generation Activity Theory. A shared object, primary outcome, and secondary outcome are identified out of the combination of the RET activity system associated with the research experience and the teacher activity system based upon the classroom, school, and district environment.

authentic research activities as prescribed by the *Framework* and *NGSS*; however, if those activities are not equally well-supported (i.e., practical) within the school activity system, the likelihood of their implementation is smaller.

This theoretical framework was further developed and chosen over others because it helps account for contextual factors that may change outcomes of the research experience or how teachers implement what they had intended. Teachers often leave their research experience with descriptions of how they will implement that are not realized due to school and other factors. This framework was chosen to help understand how these many factors might impact implementation. The framework acknowledges that the RET experience cannot lead to student outcomes independent of the school activity system in which early career teachers are interacting and growing as educators. The research study looked for student outcomes resulting from both activity systems and teacher interviews were used to characterize elements of each activity system that may have influenced those outcomes. The conceptual framework also gave us a differentiated activity system through which to compare teachers with research experience to those without since we tried to match the school activity systems as closely as possible.

1.4 | Evolution of, and Rational For, Third Generation Activity Theory

Because our efforts centered around classroom implications of engagement by RET teachers in a community of practice, our original concept was to design our study using theories surrounding Situated Learning and Communities of Practice (Lave and Wenger 1991; Wenger 1998). We wanted to compare data from the students of RET teachers with those of non-RET teachers. Adding complexity, we were also interested in looking separately at findings within subgroups, for example teachers who had also been Noyce Scholars.

The premise was that bringing teachers into direct involvement with authentic science research would not only familiarize them with the culture and practices of scientists, but also infuse them with the culture, tools, and methods of the science research community. If this was successful, we also hoped that participants would carry forward this mindset into their instructional practice and become "brokers" able to make connections across the community of schools with the community of scientific research (Wenger 1998). To assess whether our goal had been achieved, we decided to examine the current instructional practices of RET teachers and non-RET teachers. Additionally, we sought to investigate how these practices may or may not have impacted students' attitudes, thinking, engagement, and ultimately academic achievement (as evidenced by test scores). To achieve this, we decided to take a road not typically navigated and sought to gather data directly from the students of these teachers and additionally from the school districts in which records of student achievement resided.

We first ran into challenges when realizing that each of these school districts had their own system, policies, rules, and access. This affected our research methods and the data we could collect. This also meant that the rules, tools, and activity systems of these other cultures became a variable across our data sets. Findings based on data from such a complex and wide ranging study, which drew data from multiple activity systems, raised additional questions, and we realized that we needed a new way to look at these findings. Recognizing that we drew data from multiple activity systems, we turned to Third Generation Activity Theory, which addresses the idea of the complexities that arise when analyzing data from projects that integrate two or more activity systems and the conflicts that ultimately may arise from this (Engeström 2015; Engeström 2016; Foot 2001; Wade-Jaimes, Cohen, and Calandra 2019*)*. During this reflective period we began to more deeply understand how our findings were influenced by the intersecting of activity systems. It was at this juncture, that we began to create diagrams for our RET program that utilized ideas and graphic representations from Third Generation Activity Theory. As we refined Figure 1, we became more and more convinced that this theory was a fit for our research initiative.

1.5 | Purpose of the Study and Research Questions

This study seeks to address some of the challenges outlined above and extends the existing literature by investigating student achievement and student and teacher perceptions using more than self-report data and through inclusion of comparison groups. Importantly, our focus is how experiences in an RET transfer to outcomes in the classroom. While other investigations into, for example, teacher STEM identity, retention, and self-efficacy (Krim et al. 2019; Sadler et al. 2010; Avraamidou 2014) are also important in the field of understanding RET experiences, the emphasis of this investigation was on student outcomes, based both on standardized student achievement measures and through reported student perceptions of classroom interactions. As such, our research measures focus specifically on teachers and students, and not on other potential stakeholder groups such as the mentor scientists, school administrators, and overall school culture. Specifically, our study investigates the following research questions:

- *RQ1: Student Achievement*—Compared with other teachers at their schools and in their districts, are RET teachers more effective at increasing student achievement, as measured by state high stakes tests in science and mathematics, particularly in high need settings?
- *RQ2: Classroom Outcomes: Student and Teacher Perspectives*—Comparing RET and non-RET classrooms, what are differences in teacher classroom descriptions and student perceptions around STEM classroom practices, engagement, persistence in STEM course tasks, STEM career awareness, and the value of learning STEM subjects?

The study compares both classroom characteristics and impacts for the students of RET teachers with non-RET comparison group teachers and students. For the purposes of this study, participants were drawn from a pool of STAR alumni who participated in the program from 2010–2016 and are hereafter referred to as *RET teachers*. Teachers included in the comparison group for this study are referred to as *non-RET teachers*.

1.6 | Positionality of Authors

The research team is made up of professional scientists, teacher educators, and professional education researchers. Some individuals on the research team (Donnelly-Hermosillo, Horvath, Keller, Sessoms, and Vokos) participated in implementing the RET studied here, as well as other research experiences for teachers, and thus have positive expectations for these types of experiences. Another (Buxner) has served as an external evaluator for the program and other RET programs, for almost a decade. The remaining authors, who designed the study of student achievement and student outcomes and analyzed student data, are independent of the RET program development and implementation; these authors also brought analytic frameworks around STEM education more broadly.

Additionally, the positionality of the team is towards the value of science in society. While some of the team members were educators and clinical practitioners, other members are methodologists and education researchers. Members of the research team are from the US as well as other countries and bring a diversity of experiences and perspectives. Several of the authors are active researchers who publish on topics of social justice, diversity, equity, or inclusion; others are faculty active in and contributing to programs at their institutions around these areas as well. Through its support of Noyce Scholars who commit to teaching in high need schools, the RET in this study has been committed to supporting teachers and students in high need school districts, and currently also supports teacher candidates through partnerships with Minority Serving Institutions (MSIs) across the country. An important reason for conducting the study was to learn more about best practices in supporting students with emphasis on districts that have fewer resources.

2 | Methods

To address the research questions above, we collected and analyzed four distinct data sources. For RQ1, this included student achievement data from RET teachers and comparison non-RET teachers, utilizing district data from select districts that agreed to participate in this study, hereafter referred to as partner districts. For RQ2 on classroom outcomes, we collected and analyzed: (a) surveys of students of select RET teachers and matched comparison non-RET teachers; (b) interviews with select RET teachers and matched comparison non-RET teachers; and (c) a written teacher survey.

Figure 2 below highlights the number of RET and non-RET teachers involved across the study. The top row shows the data sources collected for a total of 12 RET teachers. Data was collected from five teachers across both research questions—student achievement data (RQ1) along with student surveys, teacher interviews, and teacher surveys (RQ2). Three additional teachers provided student surveys and teacher interviews, but we were not able to obtain student achievement data from their districts. Finally, student achievement data was collected from four RET teachers who did not participate in the classroom outcomes study. Student achievement data for the nine RET teachers were compared to student achievement data from a large number of non-RET teachers in the participating five districts. The specific number of non-RET teachers is unknown because data provided by partner districts was specific to students and did not contain teacher-level data. Meanwhile, classroom outcomes study data (student surveys and teacher interviews) for the eight RET teachers was compared to eight matched non-RET comparison teachers in their schools or districts.

With regard to the timing of data collection, student achievement data (RQ1) was obtained through collaboration with each partner district between Spring 2019 and Summer 2020. For RQ2, students completed surveys at the beginning of the Fall 2018 semester and again at the end of Spring 2019, and teacher phone interviews were conducted during Summer and Fall 2018 and Spring 2019, with interviewees also completing written surveys during Winter and Spring 2018 or during Spring 2019.

2.1 | Research Question 1: Student Achievement on High-Stakes State Tests

2.1.1 | Measures of Student Achievement

To answer RQ1, we employed a quasi-experimental, matching design to examine the effect of RET teachers on student achievement. This included students from nine RET teachers in comparison with matched students taught by non-RET teachers in the same districts (Figure 2). We examined student achievement in high school using the California Assessment of Student Performance and Progress System. Specifically, we used the Smarter Balanced Assessment System for math achievement and the California Science Test (CAST) for science achievement.

While the research team initially debated the usefulness of standardized assessments for measuring potentially subtle effects more strongly coupled with the RET intervention, the scope of the project did not allow for creation and validation of a customized student achievement instrument. Also, given access to student scores through our district partnerships, the use of standardized assessments was viewed as an important first step in the research, as any measurable results would have stronger policy impacts than findings obtained from a customized instrument.

2.1.2 | Sample for Student Achievement

The initial student sample included 1632 students (569 students taught by nine RET teachers and 1063 students taught by non-RET teachers)[1] from five districts who had available achievement scores in Grade 11 in 2018–2019. Note that, while hereafter we refer to these students as RET students and non-RET students, it was the teachers of these students who either did or did not participate in an RET experience rather than the students themselves. Although the RET program in this study trains teachers in both science and math, the sample of student achievement data for RET teachers that was received from

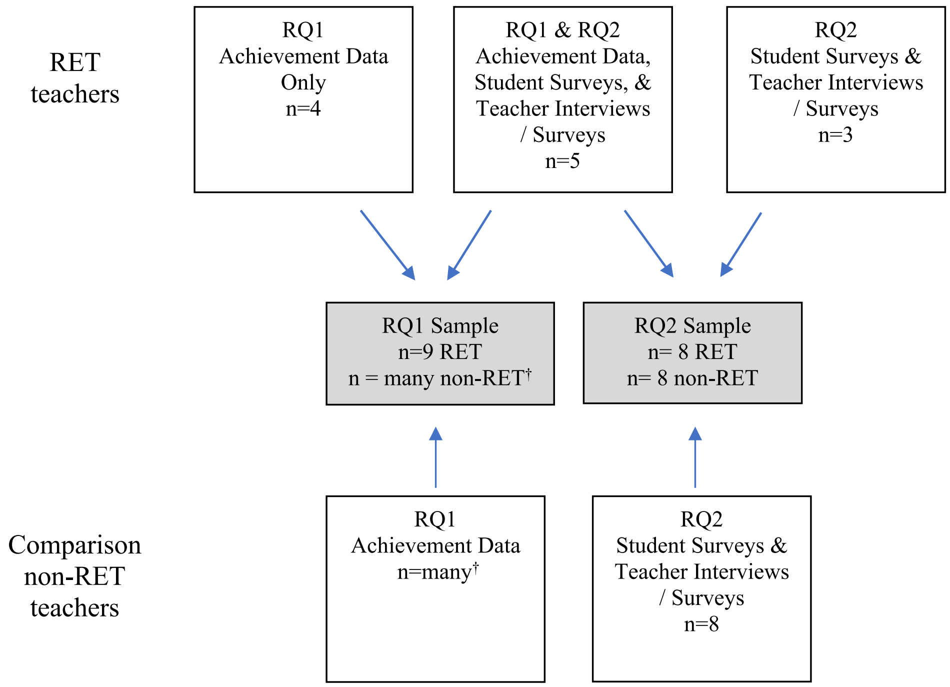

FIGURE 2 | Summary of RET and non-RET participants. *Source:* This figure presents the data associated with the 12 RET teachers who participated in the study. Data from five of the participants were used for both research questions. For research question 1, student achievement data was analyzed for nine RET teacher in comparison with students from multiple teachers in the same school and district. For RQ2, eight RET and eight non-RET teachers participated. The gray boxes in the middle show totals of RET and non-RET participants input from the top and bottom rows of the figure. †The number of comparison teachers was not specified within district level data.

TABLE 3 | Number of students in the analysis of achievement data for RET teachers.

District	RET Teacher (n = 9)	Total number of students included in analysis
District 1	RET teacher (full study)	36
	RET teacher (RQ1 only)	66
District 2	RET teacher (full study)	112
	RET teacher (RQ1 only)	120
	RET teacher (RQ1 only)	129
District 3	RET teacher (full study)	7
District 4	RET teacher (full study)	64
	RET teacher (RQ1 only)	65
District 5	RET teacher (full study)	51

Note: Table showing number of students included in analysis of student achievement data broken down by participating district and whether their RET teacher was included in the full study (RQ1 and RQ2) or was included only in the student achievement study (RQ1 only) as highlighted in Figure 2.

partner districts included only science teachers. For the grade levels requested (see below), no student achievement data for our selected RET participants who taught math was provided. Thus, our analyses for RQ1 are limited to students in science classrooms. We focused specifically on Grade 11 scores in coordination with Grade 8 baseline scores because these are the grade levels in which students complete statewide assessments in the state involved in this study. We analyzed grade-level data for these students throughout their high school career, noting in which years and grades they were taught by RET teachers (2016–17: Grade 9, 2017–18: Grade 10, or 2018–2019: Grade 11), if any. Student achievement data was collected for students enrolled in the following subjects: physics, technology, biology, chemistry, and earth science.

We used propensity scores matching to identify non-RET students who were similar to RET students to serve as the comparison group.[2] Within each of the five participating districts, we employed matching models that included students' prior achievement, background characteristics, and percentage of students eligible for free-or-reduced-price lunch at the school level.[3] Prior achievement was measured using students' Grade 8 assessments, the previous grade level that students completed the state assessments.[4] Background characteristics included race/ethnicity, gender, special education status, and English learner status. Matching was conducted separately for each district and outcome subject (science and math). For each of the nine RET teachers included in RQ1, Table 3 provides the number of students included in the analysis by district.

TABLE 4 | Student achievement sample characteristics.

Student characteristic	Students of RET Fellows	Matched comparison	Total
Science achievement			
Female	50.1%	52.0%	51.1%
Nonwhite	88.2%	87.2%	87.7%
Special education	1.0%	1.7%	1.3%
English learners	1.7%	2.9%	2.3%
Average eighth grade score	0.86	0.67	0.76
Number of students	415	415	830
Number of schools	4	17	17
Number of districts	3	3	3
Math achievement			
Female	50.2%	50.2%	50.2%
Nonwhite	79.4%	79.0%	79.2%
Special education	5.2%	6.5%	5.8%
English learners	3.8%	4.7%	4.2%
Average eighth grade score	0.59	0.51	0.55
Number of students	558	558	1116
Number of schools	8	27	30
Number of districts	5	5	5

Note: Table comparing demographic data and characteristics resulting from matched comparison efforts associated with student achievement data. For students of RET teachers the number of schools represents the number of schools attended during the year they were taught by a RET teacher. For the matched comparison group, the number of schools represents the number of schools attended in 2018–2019, the postintervention outcome year.

The matching process resulted in two samples, one for each outcome subject (science and math). The matched science achievement sample included 415 RET students and 415 non-RET students across 17 schools and three districts. The matched math achievement sample included 558 RET students and 558 non-RET students across 30 schools and five districts. No duplicate matching of comparison students occurred; rather, each RET student was matched with a unique non-RET student for the outcome subject specific outcome tests. Across both samples, about half of the students were female, most of the students were nonwhite, and very few students were English language learners or had special needs. Table 4 presents the percentages and sample sizes by condition (i.e., RET or non-RET).

2.1.3 | Analyses of Student Achievement

To answer RQ1, using the matched samples of RET and non-RET students, we conducted regression analyses to estimate the relationship between being taught by an RET teacher and student achievement. We conducted analyses separately for each district and outcome subject (math and science) using the following multilevel model to account for the nested structure of the data (Raudenbush and Bryk 2002):

$$Y_{is} = \beta_0 + \beta_1 \mathrm{RET}_{is} + \beta_2 \mathrm{RET19}_{is} + \beta_3 \mathrm{RET18}_{is} + \beta_4 \mathrm{RET17}_{is} + X_{is} + \mathrm{FRPL}_s + v_s + e_{is}$$

where Y_{is} is the outcome measure (i.e., Grade 11 math or science state assessment score) for student i nested within school s; $\mathrm{RET17}_{is}$, $\mathrm{RET18}_{is}$, and $\mathrm{RET19}_{is}$ are a set of indicators for whether a student was taught by a RET alumnus in 2016–2017, 2017–2018, or 2018–2019, respectively;[5] X_{is} is a set of student-level characteristics (i.e., Grade 8 achievement score,[6] race, gender, special education status, and English learner status); and FRPL_s is the percentage of students eligible for free-or-reduced-price lunch at the school level. Random effects are included to account for the residual effects of each school (v_s) and student (e_{is}). The main parameter of interest is β_1, which represents the effect of RET teachers on student achievement outcomes. We then pooled the estimated effects across the five districts, separately by outcome subject, using meta-analysis.[7]

2.1.3.1 | Baseline Equivalence. The matching process yielded adequate balance between RET and non-RET samples on student characteristics and prior achievement. "Adequate balance" is defined as obtaining a standardized mean difference of 0.25 standard deviation units or below (US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance 2017) on the variables included in both the matching and regression analyses, particularly those related to prior achievement. Standardized mean differences ranged from −0.04 to 0.19 for prior student achievement.

2.1.3.2 | Additional Analyses. We conducted supplemental analyses to examine whether the effect of RET teachers on student achievement differs for students in high need settings, defined as schools with more than 50% of their students eligible for free-or-reduced-price lunch. Of the five districts in our sample, two districts included all high need schools

(Districts 4 and 5), two districts included all nonhigh need schools (Districts 1 and 3), and for one district (District 2), we were unable to merge in school-level free-or-reduced-price lunch percentages due to de-identification of school level data. To determine whether the effects of RET teachers on student achievement differed by high need status, we used meta-analysis to pool the estimated effects across the districts, separately for the two high need districts and for the two nonhigh need districts.

2.2 | Research Question 2: Classroom Outcomes—Student and Teacher Perspectives

To answer RQ2, we employed a quasi-experimental, matching design to examine the effect of a select group of RET teachers on student attitudes and beliefs, and we also included teacher interviews to characterize related teacher perceptions. Classrooms and teachers were chosen using a purposive sampling method where it was assumed that those chosen would be representative of the entire sample. The RET teachers who were chosen were associated with four California State University campuses that generated the most RET alumni through the program under study and, therefore, the largest potential number of RET teachers who could take part in the classroom outcomes study. Each campus PI was asked to identify school districts that employed RET teachers, and contact those districts that (a) employed larger numbers of RET teachers; and (b) were the most likely to approve the project and collaborate with the project team (based on prior collaborative history with these districts).

Once permission was given by districts, the PIs contacted RET teachers teaching in those districts to ascertain interest in participating in the study. After generating a pool of interested RET teachers, comparison teachers were also identified based on the following criteria (in order of importance): subject taught (mathematics teachers were matched with mathematics teachers; science teachers, of any science discipline, were matched with other science teachers), grade level taught (11–12th grade teachers were generally matched with 11–12th grade teachers and 9th or 10th grade teachers were matched with 9th or 10th grade teachers), and years of experience (early career teachers were matched with other early career teachers). See Table 5 below for more details of matching information for teachers in the classroom outcomes study.

All teachers involved in this classroom outcomes study completed a teacher survey. The survey asked both RET and non-RET teachers about their classroom practices and how they taught about STEM careers, factors associated with decisions regarding staying or leaving the teaching profession, the type of research experiences in which they had participated (including the RET program involved in this study), perceptions of students' STEM learning, and views of the school environment. Survey analysis included descriptive statistics comparing RET and non-RET comparison teachers.

For each study pair, we also collected surveys of students taught by these teachers and conducted phone interviews with the teachers involved. As described in more detail below, we compared student survey results from a group of students taught by eight RET teachers during the 2018–2019 school year (i.e., RET students) to a group of students taught by eight comparison teachers (i.e., non-RET students) in the same year. Students were matched using baseline survey measures and demographics. This kind of matching ensures that we compare students who are as similar as possible across observable characteristics. Also described below, phone interviews with these outcomes study teachers were used to further characterize elements of the RET and school activity systems that may be impacting student outcomes.

2.2.1 | Measures of Student Perceptions

Guided by recommendation from the CARET literature review (Krim et al. 2019), we developed a survey to measure students' perceptions of five constructs: frequency of STEM classroom practices, student engagement, student persistence in STEM course tasks, STEM career awareness, and perception of the value of learning STEM subjects. These constructs are within the domain of social–emotional outcomes defined within the What Works Clearinghouse (WWC) framework (US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance 2017). More specifically, four of the constructs (student engagement, student persistence in STEM course tasks, STEM career awareness, and perception of the value of learning STEM subjects) fall under Intrapersonal Competencies, defined as "Mental health indicators that are primarily focused inward and reflect a student's emotional status and psychological well-being, and that include internalizing behaviors and both negative and positive feelings." (US Department of Education, Institute of Education Sciences 2021, p. 6) The above constructs represent specific primary and secondary outcomes in our revised Third Generation Activity Theory framework.

The student survey was based on existing and validated scales (Hayes et al. 2016; TIMMS 2007 Assessment Frameworks 2007; Friday Institute for Educational Innovation 2012; University of Chicago 2017) that were modified with permission from the respective authors. Modifications were made to the survey items to more closely align to the RET program. Each construct scaled score was created by combining related survey items using the Rasch rating scale model (Wright and Masters 1982) in Winsteps (Linacre 2015). As part of the scaling process, we conducted psychometric analyses to assess the survey constructs for item fit and internal consistency (i.e., whether all items hold together well and measure the same construct). We also conducted a principal component analysis to examine multidimensionality (i.e., whether a set of items grouped under a single construct measures two or more concepts instead of one). Rasch reliability, a type of reliability that accounts for the range of item difficulty and the sample size, ranged from 0.68 to 0.80 and Cronbach's alpha, which is a measure of the extent to which the items assess the same construct, ranged from 0.74 to 0.86 across the five constructs. See Supporting Information S1: Section S2 for further psychometric information by construct and for the crosswalk of individual survey items to constructs.[8]

TABLE 5 | Case study teacher participants.

Teacher (Pseudonym)	Years of RET	Years of teaching experience	Noyce	High-need school[c]	School type	Number of students in study (matched)
Sebastian	2	1–3	Yes	No	Project-based learning	28
Laura	—	1–3	No	No	Project-based learning	14
Dirk	1	7–9	No	Yes	Traditional public	41
Silvio	—	4–6	Yes	Yes	Traditional public	23
Karl	1	7–9	No	No	Traditional public	121
Gene	—	7–9	No	No	Traditional public	63
Celia	2	7–9	Yes	No	Traditional public	29
Omar	—	4–6	No	No	Traditional public	11
Sai	1	4–6	Yes	Yes	College prep charter	29
Frances	—	4–6	Yes	Yes	Traditional public	13
Rachel	1	4–6	Yes	No	Project-based learning	25
Wendy	—	4–6	No	No	Project-based learning	1
Vincent[a]	1	7–9	Yes	Yes	STEM magnet	19
Eric[a]	—	7–9	No	No	Traditional public	12
Rebecca[b]	2	7–9	Yes	Yes	Traditional public	100
Marie[b]	—	10+	Yes	Yes	STEM magnet	47

Note: Snapshot of participants in case study teacher pairs. Consistent with coloring in Figure 5, cases where both teachers in a pair had the same Noyce status are highlighted in green; cases in which the Noyce status differed are highlighted using yellow for RET teachers and blue for non-RET teachers. As described in text, students of these teachers were also surveyed; number of students in the study represents numbers of students included in analysis. [ab]Two of the eight teacher pairs were between teachers from different districts. Two teachers were the sole representatives of their districts but taught different grade levels of students. Rather than pairing these two teachers together, we split one of the within-district pairs to assign each a teacher who taught more similar grade levels in a different district. [c]A school is defined as high-need if at least 50% of students qualified for free or reduced-price lunch.

We administered two versions of the student survey. The presurvey was administered at the start of the 2018–2019 school year and served as the baseline of student perceptions. On the presurvey, students were asked to respond to prompts based upon reflections of interactions in prior math or science classes. The postsurvey was administered in Spring 2019 towards the end of the academic year, and students were asked to reflect upon interactions in the science or math class in which they were currently enrolled. Student participation in the study was completely voluntary and students did not receive any monetary or other compensation for participation. The study team followed district guidelines regarding parent consent.

2.2.2 | Sample for Student Perceptions

The initial student sample[9] contained 405 RET students taught by eight RET teachers and 412 non-RET students taught by eight non-RET teachers during the 2018–2019 school year, with an overall student survey response rate of 63% (67% for RET students and 60% for non-RET students).[10] We used a two-step approach to identify non-RET students who were similar to RET students to serve as the comparison group. First, we identified teacher pairs within districts to account for any district-level variation between RET teachers and non-RET teachers. Teacher pairs were selected based on subject taught, grade level taught, and years of experience, and made within schools if possible; if not, then within the same district. Five teacher pairs were within the same school; one pair was across two schools within the same district, and two pairs were across two school districts.

Second, we used propensity score matching to select one RET student to one non-RET student based on the closest propensity score (i.e., the comparison student with the closest propensity to being similar to students in RET teachers' classrooms) within these teacher pairs.[11] We conducted matching with replacement, allowing for non-RET students to be matched to multiple RET students. To account for this multiplicity (i.e., repeated use of non-RET students where the students are included in the data multiple times), non-RET students who were matched to more than one RET student were given less weight in the analytic model (see, e.g., Hill and Reiter 2006). The matching process identified a group of non-RET students who were most similar to RET students based on baseline scores for the outcome measures on the presurvey and background characteristics, specifically race/ethnicity, gender, student plans for highest educational attainment, and class subject (math or science). From the initial sample of non-RET students, half of non-RET students were removed because they did not achieve sufficient baseline (see subsequent section Baseline Equivalence for details).

The matched sample included 392 RET students taught by eight RET teachers and 184 non-RET students taught by eight non-RET teachers in grades 9–12. Of the sample, about half of the students were female, around two-thirds of the students were nonwhite, and most students anticipated attending a 4-year college or university or graduate school. Table 6 shows the percentages by condition (i.e., RET or non-RET), along with sample sizes.

2.2.3 | Analysis and Regression Models for Student Perceptions

Using the matched sample of RET and non-RET students, we conducted weighted linear regression analyses to estimate the relationship between a teacher's RET status and students' perceptions on the survey at the end of the school year (i.e., Spring 2019). We conducted analyses separately for each of the five constructs. We estimated the relationship between teacher RET status and student construct scores using the following model:

$$Y_{ij} = \beta_0 + \beta_1 RET_{ij} + \beta_2 Pre_{ij} + X_{ij} + Pair_j + e_i$$

where Y_{ij} is the Spring 2019, postsurvey standardized scale score for the construct of interest for student i nested within teacher pair j; RET_{ij} is an indicator of whether a student was taught by an RET alumnus in the 2018–2019 school year; Pre_{ij} is the Fall 2018 survey scale score for the construct of interest; X_{ij} represents a set of student-level characteristics (i.e., gender, race/ethnicity, grade level, highest educational attainment anticipated, an indicator of having been taught by the same teacher in the prior year, and class subject); $Pair_j$ is a teacher pair fixed effect to ensure students are only compared within the previously identified and matched teacher pairs; and e_i is the residual effect for each student. Weights were included to account for the repeated use of non-RET students (see Sample for Student Perceptions section above).

2.2.3.1 | Baseline Equivalence. We achieved adequate balance (as described previously) between the matched RET and non-RET samples on student race/ethnicity and on the Fall 2018 scale scores for the five social-emotional outcome constructs. Baseline equivalence is important because it shows that the two groups of students (RET and non-RET) are similar across student demographic information and baseline survey scale measures, which gauge their perceptions from the prior school year (i.e., before having the study teacher). Standardized mean differences ranged from −0.10 to 0.09 (see Supporting Information S1: Section S3 for more information).

2.2.3.2 | Additional Analyses. To gain a more nuanced view of how student perceptions may vary by their grade level, as well as by their eligibility for free-or-reduced-price lunch or the Noyce/non-Noyce status of their teachers, we conducted two additional analyses. First, we ran the main model with the addition of an interaction term for RET status and grade level so we could examine whether there were any differential effects based on grade level. Second, we conducted subgroup analyses based on two factors, school high need status and teacher Noyce Scholar status, to examine whether we could detect differences between RET students and non-RET students, within these groups (e.g., RET Noyce students vs. non-RET Noyce students). We focused on school high need status (with a school defined as high need if at least 50% of students qualified for free-or-reduced-price lunch) because of the Noyce program's focus on enhancing the number and preparation of teachers in high need areas. We also examined findings by teacher Noyce Scholar status because the Noyce program, similar to the RET program, provides participants with professional development opportunities aimed at improving instruction. Thus, with the Noyce

TABLE 6 | Student survey sample characteristics.

Student characteristics	Students of RET teachers	Matched comparison	Total
Female	50.5%	52.6%	51.2%
Nonwhite	64.0%	67.9%	65.3%
Taught by teacher previously	13.5%	8.7%	12.0%
Subject			
Math	7.4%	7.4%	7.4%
STEM	6.4%	0.0%	4.3%
Science	86.2%	92.6%	88.3%
Grade			
9	49.2%	8.7%	36.3%
10	11.5%	42.6%	21.4%
11	16.6%	28.8%	20.5%
12	22.7%	19.9%	21.8%
Highest education			
High school or less	5.9%	5.1%	5.6%
Career/technical school or 2-year community college	6.6%	3.8%	5.7%
4-year college or university	46.7%	50.5%	47.9%
Graduate school: MD/PhD/MA/MBA	39.3%	38.3%	39.0%
Number of students	392	184	576
Number of teachers	8	8	16
Number of schools	7	6	8

Note: Table comparing demographic data and characteristics resulting from matched comparison efforts associated with student survey data.

subgroup, we aimed to further isolate the possible impact of the RET program. We ran separate regression models for each of the five constructs, and the models for school high need status and teacher Noyce Scholar status were identical to the main model; however, we removed the teacher pair fixed effect since teacher pairs may not be intact in the subgroups (e.g., within a teacher pair, the RET teacher may be a Noyce Scholar while the comparison teacher is not) and restricted the sample for these models to each specific subgroup of students. Subsetting students to a small sample (e.g., examining responses of students in a high need school) allows us to understand how responses of students of RET and non-RET teachers may change under different circumstances and whether there are any meaningful trends within a particular subgroup.

2.2.4 | Measures for Teacher Perceptions

Individual semi-structured phone interviews were conducted with seven of the RET teachers (one teacher opted to not participate in an interview) and eight of the comparison non-RET teachers who had their students complete a student survey during the 2018–2019 school year. The interview protocol was developed based on the results of a previous study of RET teachers (Buxner 2011) and refined based on the results of the teacher survey. The prior interviews of RET teachers, supported by HHMI and the Noyce Foundation, were a pilot of the interviews used in this study that investigated how teachers talked about their own research experiences and how those knowledge, skills, and ideas were implemented in their classrooms and the barriers that they faced in doing so. The presence of contextual factors helped frame the interview protocol of this study. The full protocol can be found in Supporting Information S1: Section S4. Teachers were asked about their teaching philosophy, the RET program involved in this study and/or other research or industry experience, how they taught students about STEM careers, leadership roles they had in their schools and districts, and their classroom practices. Interviews lasted from 30 min to 80 min in length and were on average 45 min long. All interviews were audio-recorded and transcribed for analysis.

2.2.5 | Analyses of Teacher Perceptions

The interview team meeting after each interview was conducted for reflection on participant responses and on questions being asked. Analysis of interviews began with code development and revision. The first round included three members of the research team coding five of the teacher transcripts using Dedoose^{R}. Initial codes were derived from previous RET teachers' responses to an open-ended survey about how their classroom practices were informed by research experiences (Buxner 2011). The coding team met once a week while coding transcripts to discuss new codes and negotiate codes based on transcripts. The inter-rater reliability (IRR) was established by extracting relevant coded passages and having coders blindly code with the codebook so that each researcher coded the five interviews with the final codes (Salmona, Lieber, and

Kaczynski 2020). The overall resulting Cohen's kappa (k) was 0.85, with over 0.8 showing strong agreement in coding (McHugh 2012). Subsequently, the rest of the teachers' transcripts were coded by one member of the research team. Two members of the original research team (one who coded all responses and their research advisor) met weekly to review all transcripts and final coding together. An additional round of analysis was conducted with a fourth member of the research team to corroborate themes and findings. Examples of relevant codes can be found in Supporting Information S1: Section S5.

3 | Findings

3.1 | RQ1: Findings for Student Achievement

Overall, students being taught by an RET teacher in high school science courses did not lead to improved or decreased outcomes in science or math achievement, compared to outcomes for students taught by a matched comparison non-RET science teacher. Further discussion of this lack of signal in student achievement is provided in the Discussion section below.

3.1.1 | Results by District

The results generally did not differ by district. No significant achievement differences were found for science or math achievement between districts. As seen in Figure 3, all confidence intervals for science and math include zero.

3.1.2 | Results by High Need Status

We also combined results by high need status by pooling the estimated effects using meta-analysis, separately, for high need districts (Districts 4 and 5) and nonhigh need districts (Districts 1 and 3). Districts with high need schools and districts with nonhigh need schools did not differ on any student achievement outcomes. The results were not significant for either group across the subjects measured.

3.2 | RQ2: Perceptions About Classrooms

3.2.1 | Findings for Student Perceptions

Overall, based on the five constructs within the student survey, having an RET teacher was associated with more positive student survey outcomes compared with the non-RET group on three of the five student constructs: STEM career awareness (0.220 standard deviations), value of learning STEM subjects (0.199 standard deviations), and student persistence in STEM course tasks (0.291 standard deviations). Meanwhile, there was no overall statistically significant difference for student

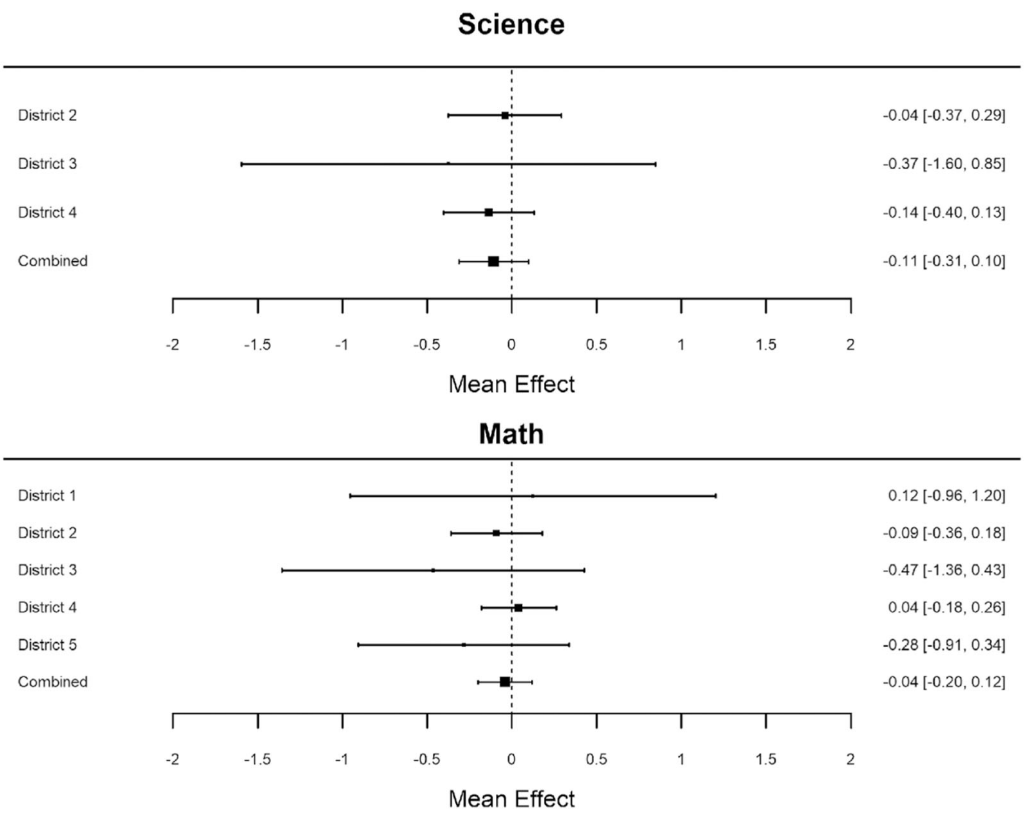

FIGURE 3 | Estimated impact of RET teachers on student achievement by district. *Source:* This figure presents the mean impact estimates for each district (or combined across districts), represented by the square in the middle of each line. Larger squares represent more precise estimates (i.e., inverse of the standard error), which also come from larger sample sizes. The value is also presented on the right-hand side of the plot. Mean effects are nonsignificant if the confidence interval, represented by the edges of the lines or the values in parentheses, includes zero; and significant if the confidence interval does not include zero.

TABLE 7 | Student survey results.

Variable	Classroom practices		Student engagement		STEM career awareness		Perception of value of learning STEM subjects		Student persistence in STEM course tasks	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
RET	−0.204	0.140	0.147	0.096	0.220*	0.094	0.199*	0.094	0.291**	0.089
Observations	576		576		574		573		574	

Note: Results from linear regression analyses of student survey data. Each column represents a separate regression. Coefficients are in standard deviation units. Student covariates are the baseline survey score, race/ethnicity, gender, subject, grade, perceived educational attainment, teacher pair, and dosage. Coefficients are statistically significant at the *5%, and **1% levels.

responses by treatment on classroom practices and student engagement. Table 7 summarizes these findings.

3.2.1.1 | Student Survey Findings by Grade Level. We conducted grade-level analyses with the full sample of students, interacting the continuous variable grade with RET status to examine whether student perceptions differ by grade level. Only one construct, the perception of the frequency of STEM classroom practices, is significant and positive (0.284 standard deviations; see Supporting Information S1: Section S6 for statistics for all constructs). Specifically, RET students in higher grades report higher frequency of classroom practices, while the reverse is true in lower grades where RET students report lower frequency of classroom practices than students of comparison teachers (Figure 4). Of the remaining constructs, general trends showing higher z-scores at upper grade levels amongst students of RET teachers are noted, though only grade-level variations in classroom practices were statistically significant.

3.2.1.2 | Student Survey Findings by School Type and Noyce Teacher Status. We also conducted subgroup analyses by restricting the sample to four subgroups of interest: students in high need schools (seven teachers), students in nonhigh need schools (nine teachers), students taught by Noyce teachers (nine teachers), students taught by non-Noyce teachers (seven teachers). We then examined whether differences exist between RET students and comparison students, within each subgroup.

As shown in Table 8, for students in nonhigh need schools, we found that students of RET teachers reported significantly more positive perceptions of STEM career awareness (0.330 standard deviations) than students of non-RET comparison teachers. In high need schools, students of RET teachers reported lower frequencies of STEM classroom practices (−0.613 standard deviations) but more positive persistence in STEM course tasks (0.419 standard deviations) than students of non-RET comparison teachers.

We also found that among teachers who were not Noyce Scholars, students of RET teachers had more positive perceptions of STEM career awareness (0.237 standard deviations) and student persistence in STEM course tasks (0.312 standard deviations) compared to students of non-RET teachers. Among teachers who were Noyce Scholars, students of RET teachers had more positive perceptions of the value of learning STEM subjects (0.644 standard deviations) when compared to students of non-RET teachers.

3.2.2 | Findings for Teacher Perceptions

In response to the written survey questions related to classroom practices, both RET ($n = 8$) and comparison non-RET ($n = 8$) teachers reported using hands-on interactive activities, promoting collaboration in the classroom, and contextualizing classroom activities around real world situations at the same rate when asked to report how often they used these practices. Comparison non-RET teachers reported using open-ended questions to stimulate whole class discussions more often than RET teachers. RET teachers reported implementing project-based learning slightly more often than comparison teachers and reported engaging students in off-campus activities more often than comparison teachers. Given the small sample size, none of the differences in survey findings were statistically significant.

When asked on the teacher survey about student engagement in SEPs, both RET teachers and comparison non-RET teachers reported having students: (a) generate questions or predictions to explore; (b) analyze results using basic calculations; (c) consider alternative explanations; and (d) seek evidence to support a claim or explanation. For each practice, comparison non-RET teachers reported engaging students more often in these practices, although the differences were not statistically different.

Analysis of teachers' interviews related to their classroom practices showed differences in how teachers talked about daily classroom practices. Overall, RET teachers described engaging students in more NGSS SEPs than comparison non-RET teachers. RET teachers included the NGSS practices of asking questions ($n = 3$), developing and using models ($n = 2$), planning and carrying out investigations ($n = 4$), analyzing and interpreting data ($n = 6$), using mathematical and computational thinking ($n = 1$), constructing explanations or designing solutions ($n = 2$), and obtaining, evaluating and communicating information ($n = 6$). Additionally, all seven RET teachers described inquiry-based practices in their classroom. The non-RET comparison teachers discussed NGSS SEPs of asking questions ($n = 1$), developing and using models ($n = 2$), planning and carrying out investigations ($n = 6$), analyzing and interpreting data ($n = 4$), using mathematical and computational thinking ($n = 1$), constructing explanations or designing solutions ($n = 1$), and obtaining, evaluating and communicating information ($n = 4$). Additionally, how these NGSS practices were discussed differed between RET and non-RET teachers. Overall, RET teachers' descriptions of NGSS practices were

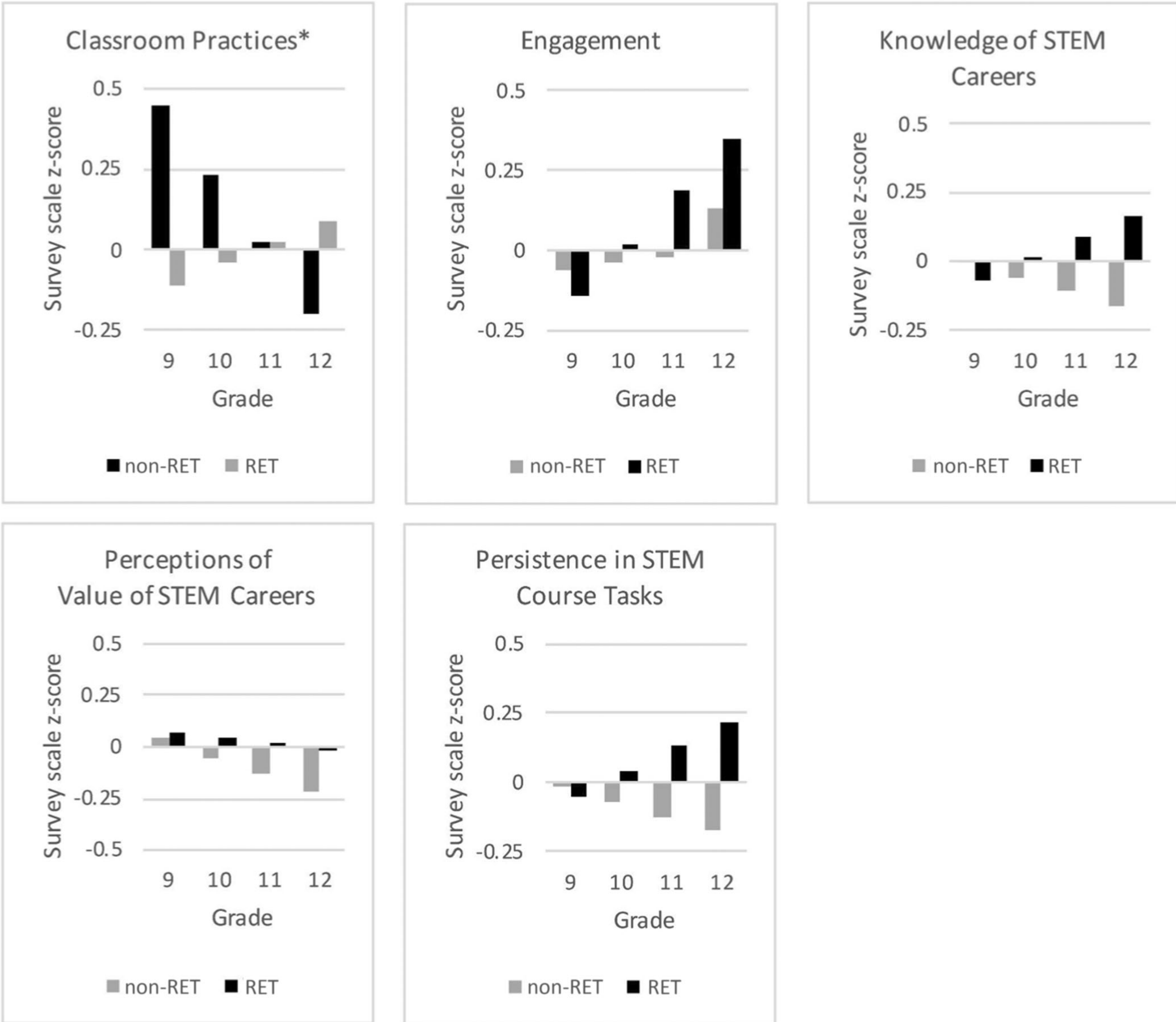

FIGURE 4 | Understanding the relationship between survey outcomes and grade. *Source:* This figure presents survey scale z-scores for each construct by grade level. The interaction between RET status and grade level was significant for only one construct—the perception of the frequency of STEM classroom practices (indicated with an asterisk above; * indicates statistical significance at the 5%). Trends are seen in all constructs favoring student perceptions of RET teachers at higher grade levels and favoring non-RET teachers at lower grade levels.

more aligned with research practices and non-RET teachers' descriptions were more aligned with structured step-by-step scientific labs. For example, in discussing obtaining, evaluating, and communicating information, one RET teacher described how his students report a lab, "The labs take a week to complete. And so they have to write abstracts for their reports. They have to come up with the extent of the design. They have to identify their control. They have to write conclusions, error analysis, all those types of things." This quote can be compared to a teacher who had not completed a research experience discussing labs, where they describe parts of a more confirmatory lab.

> *"We explored ... to determine the speed of light using a microwave and marshmallows... But you know, I would pose to them, you know, 'Okay, well this is still light in a vacuum, and give me the wave formula, okay?' and 'Okay, where ... if you could calculate the wavelength of a wave and if we knew the frequency, could you then determine what the velocity of that electromagnetic wave would be?' Okay? So then I explained to them, 'You know if you take out the roller or take out the table on a microwave, then you're gonna get a standing wave that's gonna be ... you know, the peaks and the troughs have a little more energy, so it's gonna puff up the marshmallow right in those spots.' 'And then you can plug in the formula,' and I wanna see them carry the calculations through."*

The coding group noted that RET teacher responses regarding engagement in NGSS practices were generally more sophisticated and in alignment with elements of conducting authentic

TABLE 8 | Student survey results subgrouped by school type and Noyce teacher status.

	High need school status				Noyce scholar status			
	Not high need		High need		Not Noyce		Noyce	
Outcome measure	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Classroom practices	0.147	0.194	−0.613**	0.185	0.0774	0.166	−0.353	0.212
Student engagement	0.301	0.156	−0.232	0.13	0.249	0.133	0.0928	0.145
STEM career awareness	0.330**	0.125	0.0208	0.157	0.237*	0.117	0.373	0.209
Perception of value of learning STEM subjects	0.0687	0.136	0.242	0.14	−0.0805	0.117	0.644**	0.15
Student persistence in STEM course tasks	0.193	0.122	0.419**	0.15	0.312**	0.113	0.136	0.158
Observations	303		270		262		311	

Note: Results from linear regression analyses of student survey data by subgroups. Each cell represents the coefficient or standard error for the treatment indicator in a separate model. The number of observations included in a model may vary by one or two for each outcome measure within a subgroup. The number of observations presented for each subgroup represents the smallest sample included in any of the outcome measure models for that subgroup. Coefficients are in standard deviation units. Student covariates are the baseline survey score, race/ethnicity, gender, subject, grade, perceived educational attainment and dosage. Coefficients are statistically significant at the *5%, and **1% levels. *$p < 0.05$. **$p < 0.01$.

research investigations. In their interviews, RET teachers discussed how their research experience prepared them to engage students in more open-inquiry experiences where the answer may not be known. As Rachel shared,

> *"So I guess the [RET] experience made me feel much more confident as a researcher, that I could tackle larger problems and be more.... able to be vulnerable in front of a group where you're presenting your knowledge, like I stand in front of students every day and have to have that deep confidence in myself that I might not know the answers, but I can figure out how to find them, and ... you know, because I have my students going off in so many different directions, I have no idea where they're going sometimes. And I think the [RET program] allowed or helped me feel more confident standing in front of people and not knowing the answers and feeling confident that I could find them."*

On the teacher surveys, RET and comparison non-RET teachers reported the same level of comfort and knowledge for where to direct students or parents to find information about STEM careers and where to find resources for teaching students about STEM careers. In their interviews, all teachers were asked and shared how they discussed careers with their students. Six of the seven RET teachers indicated that they currently talked to their students about STEM careers and one of the RET teachers reported that he did not currently talk about careers explicitly with his students, but he had in the past. All of the comparison teachers discussed how they talked to their students about STEM careers. One striking difference between interviews with RET teachers and comparison teachers was that five of the seven RET teachers and only one of the comparison teachers discussed not only preparing students for careers but emphasized that giving students a background in scientific thinking would serve them in any career and throughout their lives. An example of this thinking was shared by RET teacher Karl,

> *"I would like to prepare them for all jobs, and I think the way that you do that is get them to think critically. It's hard because they don't like doing it, but I try my hardest to get those brains hurting and have them really analyze, interpret data, and try to think critically, solve problems. And I think if they can do that well by the time they leave high school, regardless of what their interests are, they should do pretty well out in the work world."*

Only two of the seven RET teachers mentioned talking to their students about applying what they were learning in class toward a specific career while the other five RET teachers discussed preparing students for a wide range of careers and being a scientifically literate citizen and the value of science in society in general. Six of the eight comparison teachers described preparing students for specific careers as their primary purpose of career preparation. The resources that teachers drew upon to inform their talk about careers were dependent on their own STEM experience. Half of the RET teachers specifically mentioned their RET experience as informing their talk about careers and only a quarter of comparison teachers discussed a previous career impacting their talk about careers. Additionally, RET teachers discussed their research experience, mentors, or other connections as resources when talking about STEM careers whereas non-RET comparison teachers used more online resources to support students' knowledge of STEM careers.

Although the topic of productive struggle (Chen et al. 2024; Lowell 2024) was not part of the interview protocol, in analyzing the transcripts, six out of the seven RET teachers and two of the eight comparison non-RET teachers discussed the importance of students making mistakes, students persisting in finishing intellectually challenging tasks, and challenging students to work hard in class. Three of the RET teachers and one of the comparison teachers linked their reflections of what they desired for students back to their own research work. As Rachel reflected,

> *"Just a mindset I have. Like, I'm not worried about students making mistakes. I really try to encourage them to be pushed out of their comfort zone. And I think, again*

the one thing that I saw in that [RET] experience and that I experienced myself was it's okay to be wrong; you should just know why you were wrong. And if you can't figure out why you were wrong, then that's a more interesting question, so like that everything is leading to some learning and to some stuff for us. So I guess that's the mindset that I definitely carry over, that whatever they're doing is pushing their knowledge forward."

RET teacher Sebastian shared a similar sentiment,

I would probably have to say kind of the independence and the ... like, I had to find resources and solutions myself, like if I ran into a problem....... stuff like I try to work on with students, you know, analyzing data. Just because you didn't get the results you expected doesn't mean the experiment failed or something, so we want to come back. That's the hardest thing for them to learn... is that all data has something important to learn about it, whether that's, 'Oh we messed up the steps; that's why the data looks like this,' or, 'We didn't actually have this actually work like the real world.' 'We don't know what this actually means, so our hypothesis is wrong, but we actually considered a new hypothesis or can retest that,' or, 'We just learned something actually valuable we didn't expect to find out.'"

RET teacher Vincent shared how he encourages his female students to push through,

"A lot of them want to go into engineering and STEM in general, so I try to encourage that, especially with my female students......and I try to encourage them to be strong, be brave even if they're in a classroom full of boys."

To summarize, RET teachers discussed the research they engaged students in, as well as daily teaching practices, in more sophisticated ways than non-RET comparison teachers. RET teachers tended to draw on their own research experiences in thinking about the experiences that they wanted to create for students that were more about learning from mistakes, completing intellectually challenging tasks, and working with data. Secondary analysis of the interview findings does not show strong patterns related to the subgroup analysis that revealed significant findings in the student data (e.g., high-need, not high-need, Noyce, and non-Noyce).

4 | Discussion and Implications

This study of a specific RET for preservice and early career teachers uncovered a range of findings and new questions for inquiry. Below, we summarize key findings from our research questions and implications for related RET efforts. We also discuss challenges and limitations of the study along with implications for further research.

4.1 | Reflecting on Student Achievement Findings (RQ1)

RQ1 was intended to determine if differences in student achievement could be measured similar to three previous studies of RETs that have included student achievement data. Two of these studies used customized instruments (Ragusa and Juarez 2017; Yang, Liu, and Gardella 2020) that were narrowly focused, considered exploratory, and had no comparison groups. The third study, that included a comparison group, did find a relationship between participation of teachers in a RET and an increased passing rate on the high-stakes New York State science exam in the teacher's discipline (Silverstein et al. 2009).

Similar to Ragusa and Juarez (2017) and Yang, Liu and Gardella (2020), our study showed overall pre/post gains for all students. However, unlike the findings of Silverstein et al. (2009), no statistically significant difference in student achievement was identified on science or mathematics standardized test scores (using the California Science Test and the Smarter Balanced Assessment System, respectively). Potential reasons for these competing findings are that we used standardized instruments that were more general and that the students interacted with up to three different teachers in years between assessments. While our sample size was small and we cannot therefore exclude the usefulness of California Science Test and Smarter Balanced assessments in future work, use of customized assessments of content and learning may be more fruitful based upon comparisons of this study with findings by Silverstein et al. (2009).

To our knowledge, this study is only the second study in the existing literature, after Silverstein et al. (2009), to attempt a comparison study of student achievement for RET and non-RET teachers. The Silverstein study involved students of 32 RET teacher participants and an unspecified number of non-RET teachers in the same schools and subjects. The lack of signal in our study may have been due to limited sample sizes. A core reflection from this effort is that obtaining and analyzing assessment scores is time and resource intensive and sufficient capacity is necessary to obtain large sample sizes. To make meaningful claims around student achievement, though, inclusion of comparison groups is an important direction for the field.

4.2 | Reflecting on Student Perception Findings (RQ2)

The second research question involved student and teacher perceptions regarding classroom experiences and outcomes. For students, we investigated if their perceptions of specific constructs differed based on whether they were taught by eight RET versus eight non-RET teachers. Statistically significant positive associations were found in the domains of STEM career awareness, value of learning STEM subjects, and persistence in STEM course tasks. In the context of our activity theory framework, these represent measurable secondary outcomes resulting from the intersection of both the RET system and the school system.

RET / Non-RET / RET + Non-RET	Not High Need					High Need				Subgroup Findings
District	Traditional Public	Project Based Learning	Project Based Learning	Traditional Public	Traditional Public	STEM Magnet	Traditional Public	Traditional Public / College Prep Charter	Traditional Public / STEM Magnet	
Noyce		Sebastian	Rachel	Celia		Vincent	Silvio	Sai Frances	Rebecca Marie	Value of Learning STEM**
Not Noyce	Karl Gene	Laura	Wendy	Omar	Eric		Dirk			STEM Career Awareness* Student Persistence in STEM Course Tasks**
Subgroup Findings	STEM Career Awareness**					Classroom Practices(**) Student Persistence in STEM Course Tasks**				STEM Career Awareness* Value of Learning STEM* Student Persistence in STEM Course Tasks**

FIGURE 5 | Summary of aggregate and subgroup analyses. *Source:* This figure presents RET/non-RET groupings of case study teachers broken down by whether they were teaching in high need settings and whether they were Noyce Scholars. Consistent with the coloring in Table 5, cases where both teachers had the same Noyce status are highlighted in green; for example, RET teacher Karl and non-RET comparison teacher Gene are both non-Noyce, while RET teacher Sai and comparison teacher Frances were both Noyce Scholars. Cases in which the Noyce status differed are highlighted using yellow for RET teachers and blue for non-RET teachers; for example, RET teacher Sebastian was a Noyce Scholar and comparison teacher Laura was not a Noyce Scholar, while RET teacher Dirk was not a Noyce Scholar but comparison teacher Silvio was a Noyce Scholar. Also, note that RET teacher Vincent was a Noyce Scholar at a STEM Magnet school while comparison teacher Eric was not a Noyce Scholar and taught at a Traditional Public school in a different but neighboring district. Statistically significant findings by subgroup are shown using asterisks: * indicates statistical significance at 5% and ** indicates statistical significance of 1%. Negative correlations for RET participants are indicated with parenthetical symbol (**).

Student survey responses were analyzed across all eight teacher pairs and also through subgroup analyses involving high need and Noyce status. We present Figure 5 (which summarizes results presented in Tables 5, 7, and 8) here to guide our discussion of these analyses. Classroom outcomes study pairs in which both participants shared the same Noyce status are highlighted in green. In cases where the pair did not share Noyce status, the RET participant is highlighted in yellow, and non-RET is shown in blue.

Students of the eight RET teachers reported higher levels of STEM career awareness than students of non-RET teachers. While this finding is consistent with a prior RET survey study showing impacts on student career awareness (Autenrieth, Lewis, and Butler-Purry 2018), our study further strengthens this finding on the impact of RETs on STEM career awareness by including a comparison group. Further, this finding held in subgroup analysis for nonhigh need settings (regardless of their Noyce status) and for non-Noyce teachers (independent of high need setting).

Students of the RET teachers reported more positive perceptions in the value of learning STEM, which was also found in the subgroup analysis amongst all Noyce Fellows. Prior research, without an RET component, suggests that perceptions of the value of learning a STEM subject is an important precursor to developing self-efficacy and career expectancy, and that these perceptions are influenced by perceived benefits such as income, prestige, and self-satisfaction (Blotnicky et al. 2018; Nugent et al. 2015). Given RET students reported more STEM career awareness than non-RET students, it may be expected that perceptions of the value of learning STEM is higher for RET students. Further qualitative research investigating the perceived benefits of STEM learning for students of RET teachers would be informative.

Students of RET teachers reported more persistence in STEM course tasks than their counterparts, which was also true in subgroup analyses for high need schools and for non-Noyce Fellows. Our original intention in using this construct was to understand productive struggle within STEM courses for students of RET or non-RET teachers. However, we realized post data collection that persistence, as a construct, fails to recognize the systemic issues of gender and racial inequities inherent in STEM education with female, Black, and Hispanic students often being told to persist or to be resilient (Tan et al. 2013; McGee 2016; McGee and Martin 2011). Notwithstanding issues with persistence as a construct, our findings illustrate that teachers with research experiences may better support their students in discussing the nature of productive struggle in STEM research and overcoming these challenges. As noted in the Methods section, most students were nonwhite and about half were female, while the RET and non-RET teachers were predominantly white with a mix of male and female teachers. As we only conducted surveys of students and not interviews or focus groups, we are unable to comment further on how these students may view persistence in STEM course tasks.

When broken out to characterize student perceptions amongst younger (freshman and sophomore) students compared with more experienced (junior and senior) students, we found measurably more positive responses amongst older students of RET teachers with regard to perception of the frequency of STEM classroom practices. It is also notable that this general trend favoring non-RET teachers in lower grade levels and RET teachers at higher grades exists across all five constructs. This positive correlation with grade level for RET teachers, and anticorrelation for non-RET teachers, warrants deeper investigation. An explanation may be that types of activities described by RET teachers (see below) to engage students in deeper thinking may resonate more with older students. Researchers have also stressed the importance of developing sufficient conceptual knowledge and understanding towards cultivating student interests and attitudes towards science (Zhang et al. 2022).

4.3 | Reflecting on Teacher Perception Findings (RQ2)

Although written teacher surveys showed that both RET and comparison teachers reported similar comfort and knowledge of implementing NGSS aligned classroom practices, interviews revealed differences in how RET teachers talked about NGSS-aligned classroom practices that they used and how they integrated their research experiences into both the purpose and implementation of those practices. Similar to findings by Davidson and Hughes (2018) and Thomson and Turner (2019), RET teachers more often discussed providing experiences for their students that were similar to their own research experiences and drawing upon their research in development of classroom practices aligned with NGSS. For example, RET teacher Karl discusses how his research experience influences how he engaged students in NGSS practices,

> *"[if you came to my class] I think you would see them working through a problem together trying to get to the knowledge I want them to have by asking questions and kind of struggling through that process of not knowing immediately or not being told the answer.... so three out of the four of us who lived together—were in [our PI's] lab. We were all doing our own research.... but we did help each other. [Another Fellow] helped me out in the field multiple times. I helped him out in the field, so we were always bouncing ideas off of each other. And it was a really collaborative."*

RET teachers also indicated that they leveraged their research experience in describing the value of STEM and the importance of engaging in NGSS practices to strengthen learning and applicability to career success more broadly. This finding aligns with previous literature about the impact of engaging in research and the integration of key aspects in their classrooms (Frey, Fisher, and Smith 2019; Immordino-Yang 2016).

Findings from analysis of the teacher interviews supports the findings of the student surveys. RET teachers discussed STEM careers in consistently different ways than comparison teachers and modeled and talked about persistence in STEM course tasks in ways that may have contributed to their students' positive attitudes towards persistence in STEM course tasks. The interviews revealed additional information that did not come out in written surveys and provided a richer understanding of contextual factors that impacted how teachers talked about implementing math and science practices in their classrooms.

Overall, the RET teachers talked differently about STEM career awareness that included a broad context for understanding the value of STEM in society. Additionally, RET teachers discussed their own struggles in research and how they tried to bring that mindset to their students which may have resulted in greater student persistence in STEM course tasks.

We found it useful to reflect on our findings within the framework of multiple activity systems, and how their constitutive elements influence teachers' knowledge, skills, and dispositions, as we thought about the practicality of implementing science and engineering practices in classrooms. For example, while the STEM Education workshops intentionally included discussions of the SEPs and examples of incorporating real-world research into classroom practice, the realities of the school activity system faced by first- and second-year teachers who had participated in these discussions as preservice teachers may have limited effective implementation of this type of pedagogy. While quantitative student survey responses and achievement data were used to differentiate outcomes, qualitative interviews were useful to identify the role of the relevant activity system components as we sought to understand observed outcomes. For example, teachers revealed barriers to implementation, including pressure from school expectations, student motivation and maturity, and time.

> *"We're all teaching a curriculum that we're all expected to follow. And there's a time aspect in terms of what we have....I tend to teach more freshman and sophomores, which I feel there's a maturity level there that, to really engage in authentic research, might need to be a little more present."*

> *".... letting the students kind of play with those real-world data sets I think is really important and it's not something that I do enough of yet. But it's just hard to find time for these. You're told that you have to cover this, that, and the other thinking and time is always the primary constraint."*

4.4 | Challenges and Limitations to Study

In understanding and affirming the findings of this study, it is important and instructive to identify and describe key challenges and limitations. The challenges described here encompass a range of issues including: (a) recruitment of a broader, more representative sample of RET teachers; (b) the degree to which comparison teachers had participated in other forms of research (non-RET, but research nonetheless); and (c) questions about whether the identified standardized achievement tests measure students' understanding and application of the SEPs that undergird the NGSS or the mathematics practices embedded in the Common Core State Standards.

A core limitation lies in the challenges of drawing conclusions based upon only a dozen RET teachers involved in this study. The selection of these individuals was not entirely random, as pragmatic constraints factored into the selection process. The fact that statistically significant findings emerged within a number of constructs provides promising and compelling guidance for enhanced investigation of student perceptions with larger sample sizes of RET and non-RET classrooms. Further study with a broader sample would be important for solidifying these findings. However, it is also important to note that significant effort and coordination with districts was required in working even with this small sample size, and significant resources and capacity would be needed for an enhanced study involving larger numbers of teachers, students, and district partners.

Second, qualitative interviews and surveys involving comparison teachers revealed that while none of these teachers participated in the RET program being studied, half described prior experiences that involved STEM research elements. For example, some teachers in the comparison group engaged in undergraduate research experiences as part of their undergraduate capstone (either in a lab or in the field) along with other comparison teachers who had previously worked in industry. This insight calls into question a definitive distinction between an RET teacher–researcher and a comparison teacher with no prior research experience. While the depth, breadth, and duration of research experience and expertise varied within and between the two samples, the nature of STEM teacher preparation may make it difficult to identify a comparison pool of science and math teachers who have never experienced research. This important constraint should be considered in future research design efforts. This study provides an important effort to identify RET comparison groups, but it has also revealed associated challenges. Further research could also focus on ways in which reflective elements of an RET differ from other undergraduate research experiences that may not focus attention on translation of the research experience into classroom teaching practice.

A final limitation is whether the standardized assessments used in this study have the capacity to reveal the types of measurable outcomes and impacts on student understanding of science and the nature of science that one might expect to result from RET-prepared teachers. While on-going efforts continue to improve the depth, validation, and nuance of the Smarter Balanced and California Science Test assessment instruments, these instruments still involve a certain degree of fact-based context as a component of measurement and scoring. Future research might look beyond standardized state assessments to use assessments specifically aligned with NGSS SEPs. The instruments were selected because they represent the most standardized assessments available in the current state of the field, but arguments can be made that these may still not be sensitive to outcomes that might be more strongly influenced by teacher RET participation, such as a deeper, more nuanced understanding of how science is actually practiced.

4.5 | Insights From Study

We have found statistically significant differences in student perceptions of classroom learning within a study sample of RET teachers and comparison teachers. While limited by challenges associated with acquiring district level data and student data in classrooms, we have demonstrated a generalizable approach towards studying RET impacts beyond the approach of qualitative analysis of self-report data. We have also shown the importance of including data associated with comparison teachers. Below, we provide insights regarding our research effort that we feel are useful to share with the field.

As many researchers before us, we were and continue to be able to see the value in looking at complex organizational and educational initiatives through a more systemic view (Engeström 2016; Foot 2001; Wade-Jaimes, Cohen, and Calandra 2019). The emergence of this new perspective suggests that further studies on RET need to be framed to include and address the complexity and challenges that researching such programs incorporate, along with the identification of the different activity systems, which they ultimately impact and include. Although Third Generation Activity Theory was not our original theory for grounding our research, we now see the advantages of using this theory when launching into a research project encompassed with so many intersecting activities systems.

For future researchers interested in a study of this type, it is crucial not to underestimate the length of time needed and obstacles that exist in obtaining district approval and buy-in for participation. Within each of the six districts that initially agreed to participate as our work was proposed, there were limited numbers of RET teachers in each district. Out of this pool, eight RET teachers agreed to participate rather than the 10–15 RET teachers originally planned. Here, we emphasize the value and importance of RET programs tracking and staying in regular contact with alumni to enhance the efficiency and effectiveness of recruitment into follow-up studies like the work outlined here.

Providing authentic research experiences for teachers and opportunities for reflection on future classroom practice appears to provide a platform for positively impacting their students' perceptions in areas of knowledge about STEM careers, the value of STEM, and student persistence in STEM course tasks. The increase in STEM career awareness is perhaps not surprising given the stated emphasis in the RET program on connections to STEM career pathways. Positive findings regarding the value of STEM may have roots in the fact that the majority of RET teachers in the study conducted research outside of their primary academic discipline. This diversification and expansion of experiences amongst these teachers as teacher candidates may have strengthened their abilities to emphasize the value of STEM learning. Participants also described personal growth through their RET experiences related to persistence. Student perceptions of similar growth in persistence through their classroom interactions may be related. These findings are particularly notable given that as many as 8 years had passed between the RET and the time of our study for some of our teacher subjects.

Our findings with regard to STEM career awareness is of importance when considering the goal of broadening the participation in STEM and STEM careers by students from diverse backgrounds. A first step towards a STEM career is awareness of the breadth of careers that are available to this generation of students, beyond, for example, medical doctor, nurse, or computer programmer. Positive findings in the constructs of perception of the value of learning STEM and student persistence in STEM course tasks, are additionally important for increasing the diversity of the STEM workforce. An interesting finding is the gain in student persistence in STEM course tasks. RET teachers expressed the understanding that "wrong" answers are common, acceptable, and even necessary for scientific knowledge to grow. They talked about their own experiences with facing challenges and failure in their RET, and the need to make their own students comfortable with making mistakes. The disposition to persist in the face of challenges impacts personal and professional success, regardless of career choice, and is especially important in STEM learning and STEM professions. Another significant finding of this investigation involves trends in student reporting as a function of student grade level. In the field of research on RETs, very little attention has been paid to impact on students let alone to the grade level of the students impacted. Our effort to include this points to an important area of research related to the question of whether RETs are best suited to teachers of particular grade levels. Our preliminary findings point to RETs having a stronger impact on students at higher grade levels. Further research into impacts on students at middle school and even elementary school levels could be instructive in policy decisions regarding target teacher audiences for participation in RETs.

The strongest signal with regard to student grade level was in the construct of STEM classroom practices, which refers to student perceptions of frequency of engagement in NGSS science and engineering practices as measured by Hayes et al. (2016). In the context of this study, a possible explanation could be that courses offered to older students include more opportunities for students to ask questions, think more critically, and incorporate uncertainty into analytical thinking. As a counterpoint, freshman and sophomore students adjusting to more mature scientific thinking beyond middle school science may resist the challenges of learning STEM through the lens of experimentation, uncertainty, and more open-ended thinking (Zhang et al. 2022). A younger student more familiar with learning facts and details of a particular discipline may rate open-ended, inquiry investigations less favorably. Alternatively, the measured effect may be due to higher levels of comfort in teaching advanced science content amongst RET teachers, as indicated by research into levels of comfort in teaching science amongst elementary teachers compared to secondary teachers (Davis and Smithey 2009). If the correlation with grade level and perceptions is replicable, the above hypotheses would benefit from further, more detailed qualitative investigation. Within the context of our expanded activity theory framework, the secondary outcome of student attitudes and beliefs being impacted by the primary outcome of teacher modifications to facilitation of SEPs and classroom research appears to resonate more strongly with older students than with younger high school students.

The study can also inform efforts by coordinators of RETs in working with research mentors as well as facilitation of RET integration of research experience into teaching practice. Given initial indicators, highlighting STEM career awareness and attending to promotion of persistence in STEM course tasks (e.g., through intentional programming on growth mindset) may lead to even stronger gains in student perception. The lack of differences in student perceptions of STEM classroom practices also points to the potential need for sustained support of RET participants beyond the research experience to strengthen engagement of students in classroom research and NGSS science and engineering practices. This finding also points to the importance of understanding the school activity system, which can play an even larger role in classroom implementation than the RET activity system.

Further work is warranted to better understand whether findings in this study are causal or correlative. Additionally, we support use of our revised activity theory framework in future investigations. Research into classroom impacts should recognize that school activity systems play a role in the implementation of the practices in teaching and learning that RETs promote. This framing highlights the importance of sustained support of teacher participants as they transition from the RET experience into the classroom. This study provides promise that impacts of RET participation can be measured longitudinally and that we should continue to advance our understanding of STEM student learning downstream from teacher preparation approaches.

4.6 | Directions for Future Research

Based upon our investigation efforts, we offer the following recommendations regarding directions for future research efforts. These include: (a) characterizing the nature of STEM research experiences for science and math teachers more broadly; (b) including classroom observations of teacher–student interactions to establish stronger causal links to measure student outcomes; and (c) focusing attention to impacts of teacher research experiences on teachers and students traditionally underrepresented in STEM.

First, progress could be made through a larger and more generalized study investigating the level, degree, and quality of research experiences that STEM teachers bring to the classroom. This need was identified based upon interview findings that 12 of the 16 teacher participants involved in the classroom outcomes study described some type of prior research experience. While our teacher survey instrument included a question on whether participants felt that they had participated in prior STEM research experiences, we found a mismatch between this self-report item and interpretation of teacher interview responses regarding participant prior experience. Focus group and broader efforts to characterize and provide examples of research to conduct a targeted study of the types of research that teachers engage in could benefit the field as subsequent efforts could use this characterization to better discern the prior research experience of study participants. This type of study would also contribute meaningful insights in the field of teacher preparation more broadly, as prospectively more teacher candidates may participate in course-based undergraduate research than mentorship-style research.

A second recommendation involves the inclusion of classroom observations as an additional measure of teacher effectiveness, student–teacher interactions, and STEM learning amongst students. While our proxy of surveying student perceptions of classroom learning provided the benefit of using the student as a unit of integrated classroom experience, adding in observations of live or recorded classroom interactions could be valuable in better understanding reasons and causes for these perceptions and measures of effectiveness.

This study did not focus specifically on the impact of the RET on teachers traditionally underrepresented in STEM, as recommended by a recent review of the RET literature conducted after this study was initiated (Krim et al. 2019). Toward this end, however, we have made efforts to differentiate findings grouped by high need school/district status as well as Noyce background of study participants. Further work should be done to more strongly investigate ways in which race, ethnicity, socioeconomic background, and other factors play into measures of student performance and teacher effectiveness. As noted above, investigation of student perceptions of persistence as a research skill versus a systemic inequity could further expand our findings.

Finally, this study highlights the importance of using comparison groups in making meaningful claims regarding impacts on teacher effectiveness. Both self-report and third-party data is of limited value in the absence of comparison group measures. While challenges exist with any comparison group sampling efforts, we encourage research efforts that attempt to provide comparison group data for calibration, validation, and research power.

As shown in this study, providing authentic research experiences and opportunities for reflection on the impacts on classroom practice provides a platform for impacting student perceptions in areas of knowledge about STEM careers, positive attitudes toward the value of STEM, and student persistence in STEM course tasks.

Acknowledgements

The authors would like to thank the following organizations and individuals for contributions to this effort: the California State University System (CSU); the funders of the STEM Teacher and Researcher (STAR) Program from 2010–2016 (National Science Foundation Robert Noyce Teacher Scholarship Program, Chevron Corporation, S.D. Bechtel, Jr. Foundation, Howard Hughes Medical Institute, and National Oceanic and Atmospheric Administration); partner campus credential programs that provided credential information; advisory board members who contributed to our research and publication efforts; partner school districts, teachers, and students who agreed to participate in the study. The authors would also like to recognized several individuals who played central roles in the development of the STAR Program: former STAR Directors Susan Elrod, John Keller, Bryan Rebar, Dimitri Dounas-Frazer, and Brian Paavo; STAR Program Coordinator Kaylene Wakeman who has served the program since its inception; and Joan Bissell, former director in the CSU Department of Education Preparation and Public School Program, who tirelessly championed the STAR Program from prior to its inception. Before receiving support from the National Science Foundation, influential preliminary work on this research investigation was supported by the Noyce Foundation and 100Kin10. This material is based upon work supported by the National Science Foundation under Grant Nos. 1660839, 1660810, 1660777, 1660715, 1660658, and 2140288. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Data Availability Statement

The student and teacher survey data that support the findings of this study are partially available in aggregated form from the American Institutes for Research (AIR). Student achievement data was provided by school districts involved in the study, and AIR does not retain those data per agreement with districts. Teacher interview data are partially available through coordination with the corresponding authors.

Endnotes

[1] Not all districts were able to provide teacher-level data; thus, we are only able to report on the number of RET alumni in each district.

[2] Propensity score matching is a statistical technique that estimates the predicted probability of group membership (treatment vs. control) based on observed characteristics, and then uses that predicted probability to create a comparison group similar to the treatment group on these observed characteristics (Hansen 2004).

[3] We ran the following model to implement propensity score matching: $logit(P(\text{RET})) = \eta + \Sigma\lambda\mathbf{X}$, where RET is an indicator of whether a student was taught by a RET alumnus in the 2016–2017, 2017–2018, or 2018–2019 school year (coded as 1 for RET students and 0 for non-RET students); η is the intercept; and $\mathbf{X}$ is a set of student-level characteristics (i.e., Grade 8 achievement score, race, gender, special education status, and English learner status).

[4] For one school district, no Grade 8 data was available; thus, we used students' Grade 9 advance track status (yes/no), GPA, and credits earned for matching.

[5] We did not include indicators for whether students were taught by an RET alumnus in multiple years (e.g., in both the 2016–2017 and 2017–2018 school years or in both the 2017–2018 and 2018–2019 school years) because, across all five districts, very few students were taught by RET alumni in more than 1 year.

[6] For one school district, no Grade 8 data was available; thus, we used students' Grade 9 advance track status (yes/no), GPA, and credits earned for analysis.

[7] Meta-analysis is a statistical technique that combines results from multiple effects by weighting the contribution of each estimate of the effect based on the statistical precision with which that effect was estimated. Effects that are estimated from a larger sample (i.e., the district with the largest number of participating students) are weighted more heavily in the pooled effect, allowing the larger sample to contribute more information to calculating the overall pooled effect (Lipsey and Wilson 2001).

[8] Scaled items had to meet a cutoff of 0.60 for Rasch reliability and 0.70 for Cronbach's alpha. There was a sixth construct, perceptions of the nature of science and research, that we did not analyze because the Rasch reliability was below the cutoff.

[9] The initial sample represents students who took both the presurvey and postsurvey. We then dropped 18 students (5 non-RET and 13 RET) who were missing responses to a question asking whether they had previously been taught by their current teacher and one student who was taught by an RET alumnus in the Fall 2018 survey and a non-RET alumnus in the Spring 2019 survey.

[10] We calculated response rate by dividing the number of students by the number of students enrolled in the teacher's targeted class.

[11] We ran the following model to implement the propensity score matching: $logit(P(\text{RET}))=\eta+\Sigma\lambda X$, where RET is an indicator of whether a student was taught by an RET alumnus in the 2018–2019 school year (coded as 1 for RET students and 0 for non-RET students); η is the intercept; and X is a set of student-level characteristics (i.e., subject, gender, race/ethnicity, planned educational attainment, and Fall 2018 standardized survey measures). We did not include grade level in the matching, but prioritized matching students within identified teacher pairs.

References

Amolins, M. W., C. M. Ezrailson, D. A. Pearce, A. J. Elliott, and P. F. Vitiello. 2015. "Evaluating the Effectiveness of a Laboratory-Based Professional Development Program for Science Educators." *Advances in Physiology Education* 39, no. 4: 341–351.

Autenrieth, R., C. Lewis, and K. Butler-Purry. 2018. "Enrichment Experiences in Engineering (E3) Summer Teacher Program: Analysis of Student Surveys Regarding Engineering Awareness." *Journal of STEM Education: Innovations and Research* 19, no. 4: 19–29.

Avraamidou, L. 2014. "Studying Science Teacher Identity: Current Insights and Future Research Directions." *Studies in Science Education* 50, no. 2: 145–179.

Blanchard, M. R., and V. D. Sampson. 2018. "Fostering Impactful Research Experiences for Teachers (RETs)." *EURASIA Journal of Mathematics, Science and Technology Education* 14, no. 1: 447–465. https://doi.org/10.12973/ejmste/80352.

Blotnicky, K. A., T. Franz-Odendaal, F. French, and P. Joy. 2018. "A Study of the Correlation Between STEM Career Knowledge, Mathematics Self-Efficacy, Career Interests, and Career Activities on the Likelihood of Pursuing a STEM Career Among Middle School Students." *International Journal of STEM Education* 5, no. 1: 22. https://doi.org/10.1186/s40594-018-0118-3.

Buxner, S. R. 2011. *STEM Teacher and Researcher (STAR) Program Evaluation Report*. Tucson, AZ: Author.

Chen, Y. C., M. Jordan, J. Park, and E. Starrett. 2024. "Navigating Student Uncertainty for Productive Struggle: Establishing the Importance for and Distinguishing Types, Sources, and Desirability of Scientific Uncertainties." *Science Education* 108, no. 4: 1099–1133.

Davidson, S. G., and R. Hughes. 2018. "Communities of Practice as a Framework to Explain Teachers' Experiences Within the Community of Science." *Journal of Research in Science Teaching* 55, no. 9: 1287–1312.

Davidson, S. G., L. Z. Jaber, and S. A. Southerland. 2020. "Emotions in the Doing of Science: Exploring Epistemic Affect in Elementary Teachers' Science Research Experiences." *Science Education* 104, no. 6: 1008–1040. https://doi.org/10.1002/sce.21596.

Davis, E. A., and J. Smithey. 2009. "Beginning Teachers Moving Toward Effective Elementary Science Teaching." *Science Education* 93, no. 4 (2009): 745–770.

Doyle, W. 2013. "Ecological Approaches to Classroom Management." In *Handbook of Classroom Management*, edited by C. M. Evertson and C. S. Weinstein, 107–136. UK: Routledge.

Doyle, W., and G. A. Ponder. 1977. "The Practicality Ethic in Teacher Decision-Making." *Interchange* 8, no. 3: 1–12.

Engeström, Y. 2015. *Learning by Expanding: An Activity—Theoretical Approach to Developmental Research, 2nd ed*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139814744.

Engeström, Y. 2016. *Studies in Expansive Learning: Learning What Is Not Yet There*. Cambridge: Cambridge University Press.

Erduran, S. 2017. "Cognitive and Affective Aspects in Science Education Research: Selected Papers From the ESERA 2015 Conference." In *Visualizing the Nature of Science: Beyond Textual Pieces to Holistic Images in Science Education*, edited by K. Hahl, K. Juuti, J. Lampiselkä, A. Uitto and J. Lavonen, 15–30. Dordrecht: Springer International Publishing. https://doi.org/10.1007/978-3-319-58685-4_2.

Feldman, A., K. A. Divoll, and A. Rogan-Klyve. 2013. "Becoming Researchers: The Participation of Undergraduate and Graduate Students in Scientific Research Groups." *Science Education* 97, no. 2: 218–243. https://doi.org/10.1002/sce.21051.

Feldman, A., and D. Özalp. 2019. "Science Teachers' Ability to Self-Calibrate and the Trustworthiness of Their Self-Reporting." *Journal of Science Teacher Education* 30, no. 3: 280–299. https://doi.org/10.1080/1046560X.2018.1560209.

Foot, K. A. (2001). "Cultural-Historical Activity Theory as Practical Theory: Illuminating the Development of a Conflict Monitoring Network." *Communication Theory* 11, no. 1: 56–83.

Frey, N., D. Fisher, and D. Smith. 2019. *All Learning is Social and Emotional: Helping Students Develop Essential Skills for the Classroom and Beyond*. Alexandria, VA: ASCD.

Friday Institute for Educational Innovation. 2012. *Teacher Efficacy and Attitudes Toward STEM Survey*. Raleigh, NC: Author.

Haag, S., and C. Megowan. 2015. "Next Generation Science Standards: A National Mixed-Methods Study on Teacher Readiness." *School Science and Mathematics* 115, no. 8: 416–426. https://doi.org/10.1111/ssm.12145.

Hansen, B. B. (2004). "Full Matching in an Observational Study of Coaching for the SAT." *Journal of the American Statistical Association* 99, no. 467: 609–618. https://doi.org/10.1198/016214504000000647.

Harris, K., A. Sithole, and J. Kibirige. 2017. "A Needs Assessment for the Adoption of Next Generation Science Standards (NGSS) in K-12 Education in the United States." *Journal of Education and Training Studies* 5: 54–62.

Hayes, K. N., C. S. Lee, R. DiStefano, D. O'Connor, and J. C. Seitz. 2016. "Measuring Science Instructional Practice: A Survey Tool for the Age of NGSS." *Journal of Science Teacher Education* 27, no. 2: 137–164.

Hill, J., and J. P. Reiter. 2006. "Interval Estimation for Treatment Effects Using Propensity Score Matching." *Statistics in Medicine* 25, no. 13: 2230–2256.

Immordino-Yang, M. H. 2016. "Emotion, Sociality, and the Brain's Default Mode Network: Insights for Educational Practice and Policy." *Behavioral and Brain Sciences* 3, no. 2: 211–219.

Janssen, F., H. Westbroek, W. Doyle, and J. Van Driel. 2013. "How to Make Innovations Practical." *Teachers College Record: The Voice of Scholarship in Education* 115, no. 7: 1–43.

Kang, H., M. Windschitl, D. Stroupe, and J. Thompson. 2016. "Designing, Launching, and Implementing High-Quality Learning Opportunities for Students That Advance Scientific Thinking." *Journal of Research in Science Teaching* 53, no. 9: 1316–1340.

Krim, J. S., L. E. Coté, R. S. Schwartz, et al. 2019. "Models and Impacts of Science Research Experiences: A Review of the Literature of CUREs, UREs, and TREs." *CBE—Life Sciences Education* 18, no. 4: ar65. https://doi.org/10.1187/cbe.19-03-0069.

Latour, B. 1993. *We Have Never Been Modern*. Cambridge: Harvard University Press.

Lave, J., and E. Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511815355.

Lederman, N. G., J. S. Lederman, and A. Antink. 2013. "Nature of Science and Scientific Inquiry as Contexts for the Learning of Science and Achievement of Scientific Literacy." *International Journal of Education in Mathematics, Science, and Technology* 1, no. 3: 138–147.

Leont'ev, A. N. 1978. *Activity, Consciousness, and Personality*. Englewood Cliffs, NJ: Prentice-Hall.

Linacre, J. M. 2015. *Winsteps (Version 3.91.0) [Computer software]*. Beaverton, OR: Winsteps.com.

Lipsey, M. W., and D. B. Wilson. 2001. *Practical Meta-Analysis*. Thousand Oaks, CA: Sage Publications, Inc.

Lowell, B. R. 2024. "The Student Hat in Professional Development: Building Epistemic Empathy to Support Teacher Learning." *Science Education* 108, no. 2: 581–607.

Luera, G., and K. Murray. 2016. "A Mixed Methods Approach to Determining the Impact of a Geoscience Field Research Program Upon Science Teachers' Knowledge, Beliefs, and Instructional Practices." *Journal of Geoscience Education* 64, no. 4: 303–313.

McGee, E. O. 2016. "Devalued Black and Latino Racial Identities: A By-Product of STEM College Culture?," *American Educational Research Journal* 53, no. 6: 1626–1662. https://doi.org/10.3102/0002831216676572.

McGee, E. O., and D. B. Martin. 2011. "You Would Not Believe What I Have to Go Through to Prove My Intellectual Value! Stereotype Management Among Academically Successful Black Mathematics and Engineering Students." *American Educational Research Journal* 48, no. 6: 1347–1389. https://doi.org/10.3102/0002831211423972.

McHugh, M. L. 2012. "Interrater Reliability: The Kappa Statistic." *Biochemia Medica* 22, no. 3: 276–282.

National Research Council. 2012. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: National Academies Press.

NGSS Lead States. 2013. *Next Generation Science Standards: For States, By States*. Washington, DC: National Academies Press.

Nugent, G., B. Barker, G. Welch, N. Grandgenett, C. Wu, and C. Nelson. 2015. "A Model of Factors Contributing to STEM Learning and Career Orientation." *International Journal of Science Education* 37, no. 7: 1067–1088. https://doi.org/10.1080/09500693.2015.1017863.

Ragusa, G., and C. Juarez. 2017. Board #121: A Seven-Year Study of Middle and High School Teachers Participating in Research Experiences for Teachers Programs: Exploring the Relationship Between Teacher Performance and Student Achievement. In *2017* ASEE Annual Conference & Exposition, Columbus, Ohio, June 2017. https://doi.org/10.18260/1-2--27710.

Raudenbush, S. W., and A. S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods, Vol. 1*. Thousand Oaks, CA: Sage.

Reynolds, D., N. Yazdani, and T. Manzur. 2013. "STEM High School Teaching Enhancement Through Collaborative Engineering Research on Extreme Winds." *Journal of STEM Education: Innovations and Research* 14, no. 1: 12–19.

Richardson, S. 2016. "An Overview of the NSF Robert Noyce Teacher Scholarship Program." *Bulletin of the American Physical Society* 61, no. 1. https://meetings.aps.org/Meeting/PHYSTC16/Event/274983.

Sadler, T. D., S. Burgin, L. McKinney, and L. Ponjuan. 2010. "Learning Science Through Research Apprenticeships: A Critical Review of the Literature." *Journal of Research in Science Teaching* 47, no. 3: 235–256.

Salmona, M., E. Lieber, and D. Kaczynski. 2020. *Qualitative and Mixed Methods Data Analysis Using Dedoose*. Thousand Oaks, CA: Sage.

Shanahan, M.-C., and R. Bechtel. 2020. ""We're Taking Their Brilliant Minds": Science Teacher Expertize, Meta-Discourse, and the Challenges of Teacher–Scientist Collaboration." *Science Education* 104, no. 2: 354–387. https://doi.org/10.1002/sce.21550.

Silverstein, S. C., J. Dubner, J. Miller, S. Glied, and J. D. Loike. 2009. "Teachers' Participation in Research Programs Improves Their Students' Achievement in Science." *Science* 326, no. 5951: 440–442. https://doi.org/10.1126/science.1177344.

Snitynsky, R. B., K. Rose, and J. M. Pegg. 2019. "Partnering Teachers and Scientists: Translating Carbohydrate Research Into Curriculum Resources for Secondary Science Classrooms." *Journal of Chemical Education* 96, no. 4: 685–690. https://doi.org/10.1021/acs.jchemed.8b00793.

Southerland, S. A., E. M. Granger, R. Hughes, et al. 2016. "Essential Aspects of Science Teacher Professional Development." *AERA Open* 2, no. 4: 233285841667420. https://doi.org/10.1177/2332858416674200.

Stieben, M. E., T. A. Pressley, and M. L. Matyas. 2021. "Research Experiences and Online Professional Development Increase Teachers' Preparedness and Use of Effective STEM Pedagogy." *Advances in Physiology Education* 45, no. 2: 191–206. https://doi.org/10.1152/advan.00158.2020.

Storm, K., and G. Lichtenstein. 2019. "Standford's Summer Research Program for Teachers Long-Term Outcomes Study." *The Journal of STEM Outreach* 1, no. 1: 1–9. https://doi.org/10.15695/jstem/v2i1.04.

Tan, E., A. Calabrese Barton, H. Kang, and T. O'Neill. 2013. "Desiring a Career in STEM-Related Fields: How Middle School Girls Articulate and Negotiate Identities-in-Practice in Science." *Journal of Research in Science Teaching* 50, no. 10: 1143–1179. https://doi.org/10.1002/tea.21123.

Thomson, M. M., and J. E. Turner. 2019. "The Role of Emotions in Teachers' Professional Development: Attending a Research Experience for Teachers (RET) Program." *Education Research International* 2019: 1–12. https://doi.org/10.1155/2019/5856793.

TIMSS 2007 Assessment Frameworks. 2007. International Association for the Evaluation of Educational Achievement (IEA). TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. https://nces.ed.gov/timss/pdf/2007_8th_grade_Student_Questionnaire.pdf.

University of Chicago. 2017. *5Essentials Survey*. Chicago, IL: University of Chicago Consortium on School Research.

US Department of Education, Institute of Education Sciences. 2021. What Works Clearinghouse Study Review Protocol 1.0. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/WWC_Study_Review_Protocol_4_1.pdf.

US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. 2017. What Works Clearinghouse Standards Handbook Version 4.0.

Vygotsky, L. S. 1978. "Mind In Society: The Development of Higher Psychological Processes." In *Interaction Between Learning and Development*, edited by M. Cole, V. John-Steiner, S. Scribner and E. Souberman, Cambridge, MA: Harvard University Press.

Wade-Jaimes, K., J. D. Cohen, and B. Calandra. 2019. "Mapping the Evolution of an After-School STEM Club for African American Girls Using Activity Theory." *Cultural Studies of Science Education* 14, no. 4: 981–1010. https://doi.org/10.1007/s11422-018-9886-9.

Warburton, J., G. Hademenos, A. Eilers-Guttensohn, L. Garay, and J. B. Worssam. 2019. "Inspiring the Next Generation of Polar Scientists: Classroom Extensions From Teachers With Research Experiences." *Polar Record* 55, no. 4: 207–212. https://doi.org/10.1017/S0032247419000317.

Wenger, E. 1998. "Communities of Practice: Learning as a Social System." *Systems Thinker* 9, no. 5: 2–3.

Westerlund, J. F., D. M. García, J. R. Koke, T. A. Taylor, and D. S. Mason. 2002. "Summer Scientific Research for Teachers: The Experience and Its Effect." *Journal of Science Teacher Education* 13, no. 1: 63–83. https://doi.org/10.1023/A:1015133926799.

Wright, B. D., and G. N. Masters. 1982. *Rating Scale Analysis*. Chicago, IL: Mesa Press.

Yang, Y., X. Liu, and J. A. Gardella, Jr. 2020. "Effects of a Professional Development Program on Science Teacher Knowledge and Practice, and Student Understanding of Interdisciplinary Science Concepts." *Journal of Research in Science Teaching* 57, no. 7: 1028–1057. https://doi.org/10.1002/tea.21620.

Yezierski, E. J., and D. G. Herrington. 2011. "Improving Practice With Target Inquiry: High School Chemistry Teacher Professional Development That Works." *Chemistry Education Research and Practice* 12, no. 3: 344–354.

Zhang, L., P. A. Kirschner, W. W. Cobern, and J. Sweller. 2022. "There Is an Evidence Crisis in Science Educational Policy." *Educational Psychology Review* 34, no. 2: 1157–1176. https://doi.org/10.1007/s10648-021-09646-1.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.