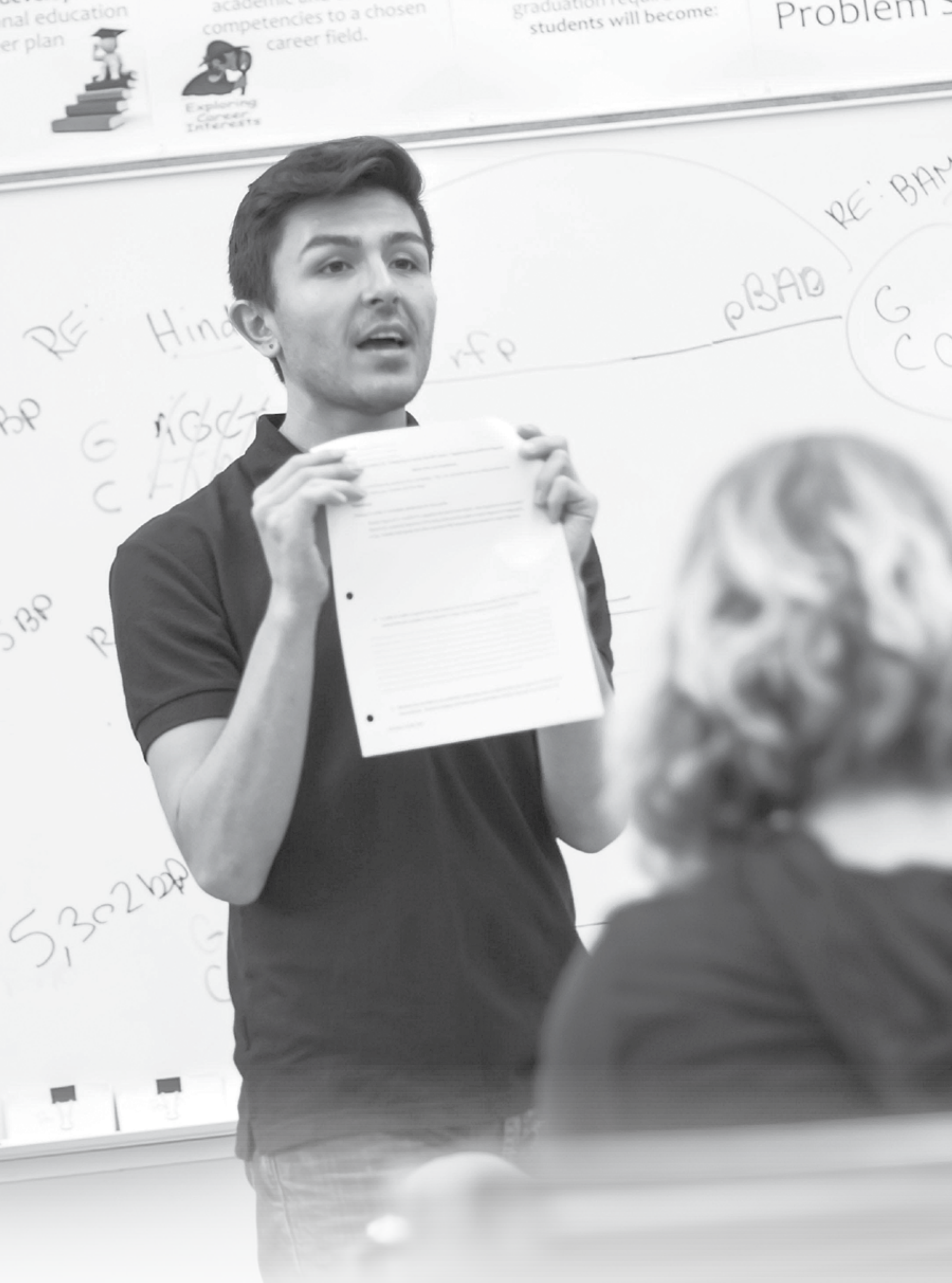
CHAPTER 4

Impacts of the STEM Teacher and Researcher (STAR) Program on Teachers, Students, and Classrooms



Michael Perez, three-time STAR fellow in 2015, 2016 and 2017 at the Santa Rosa Island Research Station [Image credit: Zach Green, Zach Green Films]

Impacts of the STEM Teacher and Researcher (STAR) Program on Teachers, Students, and Classrooms

Sanlyn Buxner, University of Arizona
Stamatis Vokos, California Polytechnic State University
John Keller, University of Colorado Boulder
Catherine Good, Baruch College, City University of New York
Dermot Donnelly-Hermosillo, California State University Fresno
Larry Horvath, San Francisco State University
Deidre Sessoms, California State University Sacramento
Elsa Bailey, Elsa Bailey Consulting
Martyna Citkowicz, American Institutes for Research
Melissa Yisak, American Institutes for Research
Dan Moreno, University of Arizona
Bo Zhu, American Institutes for Research
Eleanor Fulbeck, American Institutes for Research
Charlotte Chen, American Institutes for Research
Max Pardo, American Institutes for Research

Collaborative: NSF DUE-1660839, 1836335, 1660810, 1660777, 1660715, 1660658

> *Our findings demonstrate the value of engaging preservice educators in structured and supported research experiences to increase their productive mind-sets about themselves as participants in STEM, as well as to support their STEM-related classroom practices.*

ABSTRACT

For over a decade, the STEM Teacher and Researcher (STAR) Program has been supported by the National Science Foundation (NSF) Robert Noyce Teacher Scholarship Program along with other funders to provide paid summer research experiences for preservice and early-career science and mathematics teachers at national laboratories and other research facilities. Research into teacher research experiences (TREs) often includes self-report data from program participants, with few studies looking at impacts on students or longer-term impacts. We present results from a set of three Noyce Track 4 Research studies regarding the impact of the STAR Program on teacher effectiveness and productive mind-set. The first study compared eight STAR and non-STAR classrooms and found measurable differences both in how teachers describe their classroom practices and in how students perceive these practices and their own interests and skills, specifically in STEM career awareness, perception of the value of STEM learning, and student perseverance. The second study investigated the impact of STAR teachers on student achievement and found that there were no measurable differences in student achievement on high-stakes state assessments between students of nine STAR teachers and students of other teachers in their five respective districts. The third study found that infusing explicit instruction on productive mind-sets has a positive impact on preservice teachers' own mind-sets and the strategies that they plan to use with their students. Ongoing work involves investigating the impact of the STAR Program on teacher retention, comparing long-term retention rates of STAR and non-STAR teachers in the state of California. Our work demonstrates the value of supported research experiences for future STEM teachers and illustrates some of the complexity of assessing the impact for students.

Introduction

For decades, the model of providing mentored research opportunities during both summer and academic months has been used to prepare the future STEM workforce. The NSF has maintained a robust Research Experience for Undergraduates (REU) Program since 1987 and has also maintained a Research Experience for Teachers (RET) Program for roughly the same duration of time. Applying these models to preservice or aspiring teachers is a more recent but equally compelling approach to preparing students to learn about STEM and STEM careers.

Teacher Research Experiences (TREs), which include NSF-funded REU projects for pre-service teachers and RET projects, come in many forms. In this chapter, we focus primarily on the STEM Teacher and Researcher (STAR) Program as a model of mentored summer research for preservice STEM teachers. The STAR Program includes mentored research along with a weekly professional development education workshop and a closing conference in which participants present their research to a community of scientists, STEM educators, and teacher-researchers. The STAR Program has been funded by the NSF Robert Noyce Teacher Scholarship Program for over a decade to place preservice Noyce Scholars into 9-week summer research experiences in laboratories, primarily across the western U.S. Any Noyce Scholar, along with preservice STEM teachers from any California State University campus, can participate in STAR for up to 3 consecutive years, which can include up to 2 years of inservice participation. Robust program evaluation has been conducted since the inception of the STAR Program. More recently, we have leveraged

15 years of alumni into a set of follow-up studies of STAR teachers and have added sets of comparison teachers who have not participated in this TRE program to increase the validity of our findings.

Background

The STAR Program has conducted rigorous evaluation since its inception. Here we report on a set of three studies, funded by two Noyce Track 4 Research awards, that have been conducted to investigate the impact of the STAR Program. These studies include two comparative research studies of the impacts of the STAR experience on perceived outcomes among students and student achievement scores as measures of teacher effectiveness (DUE 1660839, 1660810, 1660777, 1660715, and 1660658). A third study investigated changes in productive mind-sets among STAR Fellows across their summer research and support workshops (DUE 1836335). These studies have been conducted through collaborations of faculty from the California State University system, University of Colorado, University of Arizona, Baruch College, CUNY, and researchers from the American Institutes for Research. These studies, including the research questions and data sources, will be described following the literature review and overview of the STAR Program.

Teacher Research Experiences

Assessment of TREs has been conducted for over 20 years (Westerlund et al., 2002). Research on TREs illustrates many positive outcomes for teachers completing these programs. Such benefits include improved self-efficacy, enhanced understandings of the nature of science and inquiry, better science content knowledge, refined or extended teaching practices, and greater knowledge of science practices (Evans & Crippen, 2020; Feldman & Ozalp, 2009; MacFadden et al., 2020; Southerland et al., 2016). However, research has also documented challenges in identifying effective research mentor-research mentee matches, as well as issues of collaborations between teachers and scientists, as scientists can downplay or undervalue teacher expertise (Shanahan & Bechtel, 2020). Teachers may also be initially intimidated by scientists' expertise in a given field when first joining a research project (Shanahan & Bechtel, 2020). Further, studies of teacher conceptual gains in TREs show mixed results (Feldman et al., 2013; Luera & Murray, 2016; Westerlund et al., 2002). As such, despite potentially positive TRE outcomes, there is still much to be learned about TRE design features that ensure mutually beneficial outcomes for scientists and teachers.

Beyond outcomes directly self-reported by TRE participants, there is a dearth of research illustrating how TREs translate into classroom practices and, in particular, how students of these teachers are impacted. Previous studies of TREs have predominantly relied on teacher accounts of student experiences and have not included students' perspectives. Further, there has been limited insight not only into student perceptions but also into student conceptual learning as a result of having a teacher with a TRE. Two studies with firsthand accounts from engineering TRE participants found that their students showed increased subject awareness and increased interest in pursuing an engineering career (Autenrieth et al., 2018; Reynolds et al., 2013). Studies on conceptual learning of students of TRE teachers showed increased pre-/post-content knowledge (Ragusa & Juarez, 2017; Yang et al., 2020). One study has shown that students of TRE teachers outperform students of non-TRE teachers in the New York Regents exam (Silverstein et al., 2009).

Overall, the existing literature illustrates the relative infancy of research into TRE programs. Many programs are still determining effective design features of the TRE for intended outcomes. Furthermore, limited studies of TRE program alumni effectively investigate how TREs translate into STEM classrooms, from teacher and student perspectives. Only one study has included comparison teachers when assessing student conceptual learning for TRE teachers (Silverstein et al., 2009). Hence, there is still much to learn about the impact of TREs.

Productive Mind-Sets

U.S. cultural assumptions about STEM ability are sometimes rooted in innate, talent-based views. These views encompass beliefs that some people are "STEM people" and some are not, and that these abilities are conferred at birth (Williams & King, 1980). Holding this "fixed view" of intelligence has been shown to lower motivation, undermine performance, and disrupt learning, especially when learners are faced with challenging content (Dweck & Leggett, 1988; Dweck & Yeager, 2019; Grant & Dweck, 2003; Mangles et al., 2006; Yeager & Dweck, 2012). Furthermore, recent research has established a strong connection between the extent that a field is imbued with a culture of talent and the lack of representation of female and Black students in that domain (Leslie et al., 2015). The alternative to this view—a "growth mind-set"—eschews these talent-based foundations for success and instead embraces the idea that intelligence can be developed over time through hard work, perseverance, and productive effort (Dweck & Leggett, 1988). In learning settings, all benefit when students are focused on improving, rather than proving, their ability, motivation, performance, and learning (Yeager et al., 2019). Further, research has shown that intentionally and explicitly teaching students about the malleability of intelligence—that is, that neurological and biological changes happen in the brain in response to novel, challenging content—not only boosts students' test scores, but also increases student achievement across the board in mathematics (Aronson, Fried, & Good, 2002; Good, Aronson, & Inzlicht, 2003).

Mind-sets about intelligence do not exist in a vacuum. Rather, the learning culture that students experience can affect their own beliefs and perspectives (see Bielock, 2010), and teachers' own fixed or growth mind-sets can impact the culture of talent in their STEM classroom. For example, teachers' mind-sets affect their pedagogical practices (Anderman et al., 2001) in ways that align with those beliefs, such as highlighting achievement of students rather than their growth and development. In addition, teachers with a fixed mind-set have been shown to be more likely to implement a performance approach rather than a learning approach in their classrooms, to lower expectations for struggling students, and to blame students for their own struggles (Rattan et al., 2012) rather than focus on improving pedagogy (Gooyabadi & Good, 2016). Students not only accurately discern their teachers' mind-sets through these differentiated pedagogical approaches but respond with reduced motivation, interest, investment in the class (Rattan et al., 2012) and grades (Canning et al., 2019) when teachers hold fixed rather than growth mind-sets. Thus, promoting growth, rather than fixed, mind-sets is important for all students. This approach is particularly important for prospective STEM teachers who will impact the mind-set of their future students (Dweck, 2016).

Promoting growth mind-sets is particularly critical within STEM domains whose culture of talent not only undermines students' foundations for success but also amplifies harmful messages that only some people truly belong in the domain (Good, Rattan, & Dweck, 2012). Sense of belonging is a fundamental human motive (Baumeister & Leary, 1995), and as such, can affect important outcomes such as physical and mental health. Feelings of belonging to an academic domain are critical (Steele, 1997), especially for learning. High academic belonging not only reflects feelings of being inside a discipline or learning community rather than on the fringes of it (Good, Rattan, & Dweck, 2012) but also predicts a suite of important variables related to learning success, such as anxiety, confidence, utility value, and intrinsic motivation.

The foundations upon which students base their feelings of belonging can also be important to consider. For example, do students feel like valued members of the STEM community because of their own efforts and engagement, or does belongingness depend on their achievement?

Emerging research (Good et al., 2015) has shown the importance of fostering effort-based belonging (growth mind-set) over achievement-based belonging (fixed mind-set) for STEM outcomes. Specifically, students whose belonging is contingent on effort (rather than achievement) not only engage more with their learning task but also are less vulnerable to both cultural stereotypes, suggesting that some people have more ability than others in STEM and to talent-based messages in their learning environment (Good et al., 2015). This is especially important for underrepresented groups in STEM. Often, teachers employ "comfort strategies" when a student does not do well to help them feel better, instead of encouraging them to try productive approaches (involving more or different kinds of effort) to succeed, aligning with a fixed mind-set for some of their students (Gooyabadi & Good, 2016). Thus, developing a hardy effort-based sense of belonging is also important for both STEM students and teachers, and especially for teachers who work with high-need students. As discipline-based education research in all STEM domains has amply illustrated, the learning of STEM inevitably involves conceptual and problem-solving struggle through engagement with tasks of high cognitive demand. If this struggle is interpreted by the student (or worse, by the teacher) as evidence that the student has reached immutable limits of their innate STEM intelligence, then the response to this struggle will be profoundly different than the response of those who recognize that everything worth learning is difficult and requires effortful practice. Normalizing productive struggle, therefore, is an important orientation for future STEM teachers. It is important to note that in recent years, misunderstandings about fixed and growth mind-sets in educational contexts have served to place the blame that properly belongs to systems on individual students (see, for instance, Good & Regnier, 2016).

Conceptual Framework: Studying TREs

Understanding the value of research experiences for undergraduate students and teachers continues to be an active area of research (Krim et al., 2019). The Collaborative Around Research Experiences for Teachers (CARET) is a group of stakeholders from U.S. research programs from across the country that has come together to discuss systematic ways to assess TREs, undergraduate research experiences, and course-based undergraduate research experiences in science courses. As part of this work, the group reviewed 307 empirical papers studying research experiences and developed a research-based model for studying research experiences (Krim et al., 2019). The CARET model conceptualizes a relationship among TRE communities of practice, professional learning communities, and outcomes for teachers and their students. The CARET model lays out the potential impacts of TREs to involve teachers' STEM research experience, knowledge, practices, awareness, attitudes, and perceptions, which in turn results in changes to classroom practices. Those practices, if enacted, are predicted to result in changes in student outcomes. In addition to the CARET framework, we draw upon a model of third-generation activity theory (Engeström, 2015; Wade-Jaimes, Cohen, & Calandra, 2019) to conceptualize interactions between TREs and classrooms, as teachers make decisions about what can be implemented in the classroom from their research experiences. Taken together, this conceptual framework helps drive the direction of data collected and an analysis framework to understand the complexity of not only how research experiences may be implemented but also barriers to implementation in classrooms.

Context for Research: STAR Program

The STAR Program is a national initiative headquartered at California Polytechnic State University, in San Luis Obispo, California. It provides immersive 8- to 10-week summer research experiences to prospective STEM teachers from any California State University campus or any Noyce Program in the nation. Noyce students must start their first STAR placement when they are participating in either a Track 1 or a Track 2 Noyce Program. To be eligible to participate in STAR, a student needs to be at least in their junior year, but not have begun teaching in their own classroom, by the beginning of the year in which they participate for the first time. Participants may return for two additional summers until they complete the second year in which they are the teacher of record.

Since its inception in 2007, STAR has served 590 aspiring and early-career participants in 820 research placements at more than 40 cooperating research settings including national labs funded by NSF, the National Aeronautics and Space Administration, the Department of Energy, the Department of Defense, the National Oceanic and Atmospheric Administration, and the United States Geologic Survey, as well as university laboratories and research and development settings. Successful STAR applicants have come from all 22 California State University (CSU) campuses with teacher education programs and 88 Noyce projects throughout the United States. STAR Fellows have impacted more than 200,000 pre- college students in their own classrooms. STAR Fellows have aspired to pursue teaching (or are in their first or second year of teaching) STEM subjects at the middle and high school levels (90%) as well as the elementary school level (10%).

The STAR experience has the following components, which together form the backbone of the STAR model:

1. A 9-week paid research experience with a mentor chosen by a lab site coordinator. Occasionally, with the mutual agreement of the STAR Fellow and their mentor, STAR extends the appointment for a tenth paid week. Similarly, the experience may be shortened by a week (with a concomitant reduction in the stipend) to accommodate a Fellow's need to assume their professional teaching responsibilities at an early-start school district.

2. A concurrent weekly 3–5-hour STEM Education Workshop, organized around geographically close face-to-face placements (pre-COVID-19) or in remote cohorts that include STAR Fellows with varying experience in STEM teaching (during COVID-19).

3. A 2-day closing research conference, which until 2018 was a STAR-specific conference. Starting in 2019, the STAR closing conference has been held in conjunction with several other TRE programs, with substantial support from the M. J. Murdock Charitable Trust through its Partners in Science National Conference.

Both the research experience and the closing conference are similar to other TRE programs. Key to the research is that participants work in STEM research labs, are mentored by active science researchers, and do independent research that contributes to the overall research program for that lab. At the closing conference, STAR Fellows come together to present their summer research findings (through poster presentations and/or oral presentations) and to network.

A central feature of the STAR TRE model, the STAR STEM Education Workshop, provides crucial support that helps Fellows translate the research placement into productive K–12 classroom practice. With specially designed experiences that bridge the laboratory with the classroom, STAR Fellows are encouraged to develop lessons and opportunities for pre-college students that will engage them in authentic science and engineering or mathematical practices, which is a major goal of the Next Generation Science Standards and the Common Core Standards in Mathematics for all students. Originally, the workshop involved two deliverables: a research poster (as an enduring, authentic disciplinary product) and a lesson plan that reflected, if not the research focus of the STAR Fellow, at least the practices of the laboratory. In addition, topics of reflection in the workshop included Nature of Science considerations, a focus on the plethora of STEM careers open to students, and professional advice about job-seeking in school systems. Since 2018, STAR has deemphasized the second deliverable, moving it from individual lesson plans to lesson plans developed by a group of STAR Fellows, and introduced an explicit focus on productive mind-sets.

Although a primary goal of STAR is for teachers to create authentic, scientific research experiences in their own classrooms, the STAR leadership reflected that a precursor to succeeding in this goal is the development of STAR Fellows' dual identities as teacher-researchers. That is, it is necessary to recognize the need to provide experiences for STAR participants that will help them build resilient identities not just as STEM teachers but also as STEM researchers. Consequently, the program turned to the literature in social psychology, described above, to identify key psychosocial variables that research has shown to be predictive of STEM outcomes, such as identification with STEM domains. The program leadership chose to focus on mind-sets about the nature of STEM intelligence (Dweck, 2013) and mind-sets about who truly belongs in STEM (see Good et al., 2012).

It is the intention of the program that explicitly discussing these mind-sets—beliefs about intelligence (e.g., fixed and growth mind-sets) as well as belonging—not only can help STAR Fellows understand that productive struggle is part of the process of research but also can aid in retaining STEM teacher-researchers. Each weekly, 3–5-hour educational workshop session includes opportunities to develop Fellows' deep understanding of growth mind-set and belonging. For example, Fellows work on a challenging mathematics problem that necessitates productive struggle and then discuss their responses to that struggle. Fellows also discuss their work in their research lab and focus on the shared experience of struggling within the content and process of doing research. These conversations are intended to normalize their experiences of challenge and help establish the universality of feeling confused or unsure in science. Research has shown that when hard work and struggle are conveyed as normal processes of academic work, feelings of belonging are boosted (Smith et al., 2012).

Investigations

Motivated to better understand the impact of the STAR Program on current Fellows and on alumni as teachers, a group of researchers designed and implemented three research studies to explore the impact of STAR on teachers' classroom practices, potential impacts on student achievement, and the impact of newer innovations in the weekly workshops on Fellows' mind-sets and feelings of belongingness. The first two studies are of classrooms of alumni who participated in STAR between 2007 and 2018 and are active science or mathematics teachers. Both studies employ a matched design and include student data. The third study, of active STAR participants, was carried out to understand how changes related to infusing a productive mind-set would impact the participants' own mind-sets and feelings of belongingness in STEM.

Study 1: Classroom Practices and Teacher Effectiveness

The purpose of this study was to determine the impact of a STAR research experience on teachers' classroom practices as a way to investigate teacher effectiveness. Data were collected from teachers and their students from classrooms of STAR teachers and, for comparison, of non-STAR teachers. The overall study was guided by the following research questions:

1. Student Perceptions—Compared to students of non-STAR teachers, how do students of STAR teachers differ in their perceptions of STEM classroom practices, engagement, perseverance, STEM career awareness, and the value of learning STEM subjects?
2. Classroom Practices—Compared to other teachers in their schools and/or districts, how do STAR teachers talk about classroom practices and student learning?

Methods for Investigating Classroom Practices and Teacher Effectiveness

This study collected data from teachers and students in 16 classrooms, eight STAR and eight non-STAR. The non-STAR comparison teachers were selected because they taught related subjects in the same school or district as the STAR teachers. Data included student perceptions as well as teachers' narratives. Students' perceptions were gathered through surveys and included constructs related to their teachers' classroom practices, and student engagement, as well as questions on their own STEM career awareness, perception of value of learning STEM subjects, and student perseverance. The study employed a quasi-experimental, matched design to examine differences in these constructs between the study sample of classrooms of STAR alumni and non-STAR matched comparison groups. The survey was administered to students during both the beginning and the end of the 2018–2019 academic year. The 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) with modifications.

During the pre-administration of the survey, a total of 817 students (405 students taught by eight STAR teachers and 412 students taught by eight non-STAR teachers) were asked to rate their prior STEM learning experiences; during the post-administration, students were asked to rate their experience with their current STEM teacher. The overall response rate was 63% (67% for students of STAR teachers and 60% for students of comparison teachers). A matched comparison analysis was conducted amongst student survey respondents. This kind of matching ensures that we compare students who are as similar as possible across observable characteristics (that is, characteristics we can measure, such as baseline scores for the measures on the pre-survey, and background characteristics such as 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). Accounting for students who did not fully complete the survey and the propensity matching resulted in a total of 392 students taught by a STAR teacher matched with 184 students taught by non-STAR teachers. **TABLE 1** below gives some of the demographics of the sample of students used for the comparison.

TABLE 1

Example Student Demographics Used for Student Matching

Student Characteristics	Students of TRE Teachers	Matched Comparison	Total
Female	50.5%	52.6%	51.2%
Non-White	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%
Number of Students	392	184	576
Number of Teachers	8	8	16
Number of Schools	7	6	8

A weighted linear regression was used to estimate the relationship between the teacher's TRE status and student perceptions for each construct. Additionally, subgroup analyses investigated both high-need status (based on school-level data) and Noyce status of each teacher.

Phone interviews were conducted with each teacher to obtain teacher narratives about their classroom practices. These interviews were recorded, transcribed, and coded by a team of three researchers. An initial codebook was developed before coding based on the constructs of the student survey. An initial coding of transcripts led to the addition of codes related to both student constructs and emergent themes. Each transcript was coded by at least two researchers. Inter-rater reliability was 0.85. All transcripts were coded with the final codebook and emergent themes were analyzed with attention to STAR and non-STAR teachers.

Classroom Practices Findings

Below we summarize the key findings of the study. More detail about analyses, full results, and findings can be found in Keller et al. (2022).

In comparing student survey responses between STAR teachers and comparison teachers, statistically significant findings were obtained in the domains of student perseverance ($p < .01$) as well as STEM career awareness and perception of values of learning STEM subjects ($p < .05$). **TABLE 2** below summarizes these differences. The regression coefficients can be found in Keller et al. (2022). Students of STAR teachers reported stronger gains in STEM career awareness, value of learning STEM subjects, and student perseverance.

TABLE 2

Student Survey Main Model Regression Summary of Result

Student Survey Construct	*n*	
Classroom practices	576	Negative coefficient
Student engagement	576	Positive coefficient
STEM career awareness	574	Positive coefficient*
Value of learning STEM subjects	573	Positive coefficient*
Student perseverance	574	Positive coefficient**

Note. Each row represents a separate regression. Coefficients are statistically significant at the * 5% and ** 1% levels.

Subgroup analysis revealed that the STEM career awareness increase for students of STAR teachers was consistent in non-high-need settings regardless of Noyce status. The same was true for non-Noyce teachers. Students of STAR teachers reported stronger gains in perceptions of the value of learning STEM in aggregate and for non-Noyce teachers. Students of STAR alumni had stronger gains in student perseverance in aggregate, for teachers in high-need schools and for non-Noyce teachers. Lastly, an analysis of students in high-need schools revealed stronger frequency of self-reported classroom practices for non-STAR teachers.

TABLE 3 below summarizes the regression coefficients for the subgroup analysis. All of the coefficients were positive except the one for classroom practices in the high-need subgroup analysis (noted in the table).

TABLE 3

Summary of Findings From Subgroup Analysis of Student Surveys

	Not High Need	High Need	Subgroup Finding for Constructs
Noyce	3 STAR teachers	3 STAR, 3 non-STAR teachers	Value of learning STEM subjects (**)
Not Noyce	1 STAR, 4 non-STAR teachers	1 STAR teacher	STEM career awareness (*); student perseverance (**)
Subgroup finding for construct	STEM career awareness (**)	Classroom practices (negative coefficient **); student perseverance (**)	STEM career awareness (*); value of learning STEM subjects (*); student perseverance (**)

Findings from the interview analysis gave some insights into potential factors contributing to student survey responses. In particular, STAR alumni noted that their efforts to promote STEM career awareness and to promote the value of learning STEM subjects were more agnostic to a particular STEM subject or topic. Rather, the STAR teachers emphasized STEM learning as having value in any career or discipline beyond high school and valued the role of scientific thinking in general. Additionally, STAR alumni were able to draw upon their own experiences doing research and their own connections for resources related to STEM careers. Comparison non-STAR teachers were more likely to talk about preparing students for specific STEM careers and relied more heavily on online resources or other sources to support career connections.

Interviews with teachers highlighted STAR teachers' empathy regarding student perseverance as a result of challenges faced during their summer research experiences. While not the only contributing factor, STAR alumni appreciation for working in an unfamiliar laboratory environment on a research problem that does not have an easily defined or scripted solution may have contributed to enhanced responses with regard to student perseverance. Interview findings also revealed differences in how STAR teachers talked about engaging their students in science and engineering practices more often in the classroom.

STAR teachers discussed how their own research informed engaging students in more open inquiry activities, their comfort with letting students explore questions for which they did not already know the answer, and the importance of problem-based activities.

Overall, the findings of this study revealed measurable differences in classroom practices described by STAR teachers and outcomes for students measured through student perceptions. The implication of this work is that engaging in research experiences with pedagogical support may have a strong impact on classroom practices, which can be detected both through conversations with teachers and through student perceptions.

Study 2: Student Achievement

The purpose of the second study was to determine if changes in teacher effectiveness found in the first study were detectable in student achievement scores. Data for this study included standardized student achievement test data collected for STAR and non-STAR teachers who taught similar grades and subjects in the same districts. Inspired by gains in performance on the New York Regents Board exam shown by students of 2-year participants in a summer research experience (Silverstein et al., 2009), this study was guided by the following research question:

Compared with other teachers at their schools and in their districts, are STAR teachers (Noyce and non-Noyce) more effective at increasing student achievement in science and mathematics, particularly in high-need settings?

Methods for Investigating Student Achievement

The project worked with administrators from five partner districts that hosted our case study teachers involved in Study 1. Each district was asked to provide student achievement data in mathematics and English language arts from the Smarter Balanced Assessment (California Department of Education, 2022) and in science (from the California Science Test, or CAST). Data were collected for students of nine STAR teachers (five of whom were also involved in data collection for Study 1) matched with students of comparison teachers from their districts. Districts provided student achievement scores from students as eighth graders and as 11th graders along with a flag about whether and how often the students had been taught by a STAR teacher during grades 9, 10, or 11.

Propensity scores matching was used to pair students of STAR science teachers with students of non-STAR science teachers in their districts. This resulted in 415 student pairs who completed the science assessment across 17 schools in three districts, 558 student pairs who completed the mathematics assessment across 30 schools in five districts, and 566 student pairs who completed the ELA assessment across 30 schools in five districts, with no duplicate matching of comparison students. Matching was based upon students' prior student achievement scores from eighth grade; background characteristics of race, gender, special education status, and English learner status; and percent of students in free or reduced-price lunch programs at the school level. Baseline equivalence was measured to ensure that standardized mean differences were lower than 0.25 standard deviation units. Across all three samples, about half of the students were female, most of the students were non-White, and very few students were English language learners or had special needs. **TABLE 4** presents the percentages and sample sizes for the science test, as an example of matching, by condition (e.g., STAR or non-STAR).

TABLE 4

Student Achievement Sample Characteristics

Student Characteristic	Students of TRE Fellows	Matched Comparison	Total
Science Achievement			
Female	50.1%	52.0%	51.1%
Non-White	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

Note. For students of TRE alumni, the number of schools represents the number of schools attended during the year they were taught by a TRE Fellow. For the matched comparison group, the number of schools represents the number of schools attended in 2018–2019, the post-intervention outcome year.

For each of the three subject areas, regression analyses were conducted using a multilevel model to determine if a correlation existed between student achievement scores and being taught by STAR alumni during 9th, 10th, or 11th grade. Effect sizes and confidence intervals were calculated for each subject area by district to determine statistical significance.

Student Achievement Findings

For all districts and in aggregate, regression analysis indicated no statistically significant differences in mathematics and science achievement when comparing students of STAR teachers and students of other teachers in the school or district. Similarly, no significant differences were found for English language arts (ELA) within four of the five districts, but we did find a statistically significant decrease in ELA scores amongst the students of three STAR alumni compared with other students in the district (effect size of -0.33 with confidence interval of 0.28). The full analyses, results, and findings can be found in Keller et al. (2022).

In summary, our efforts to investigate student achievement gains revealed no measurable differences across all outcomes and districts sampled except in a single district with regard to English language arts. We discuss these findings and implications for future research at further length in the Discussion section of the chapter.

Study 3: The Impact of Productive Mind-Sets

The third study was of active STAR Fellows, those participating in STAR actively either as preservice or early-career teachers during the 2018 and 2019 summers research programs. This study investigated the following research question:

To what extent does an explicit focus on growth mind-set and belonging lead STAR Fellows to (a) adopt the view that STEM intelligence is a malleable quality that can be increased with effort and engagement, and (b) strengthen their feelings of belonging to the STEM community?

Methods for Investigating Productive Mind-Set

Data were collected during the 2018 and 2019 summers as well as the academic years that followed. STAR Fellows (n = 103) completed a survey comprised of questions that measured Fellows' (a) mind-sets of the nature of STEM intelligence as either a fixed trait or a malleable quality ("You can learn new things in STEM, but your basic STEM intelligence stays the same," Good et al., 2012), and (b) feelings of belonging to the science community ("I feel like a member of the STEM community," Good et al., 2012). The survey items were drawn from the literature and had strong reliability (alpha = .86). On the survey, Fellows rated a series of statements on a scale from 1 to 8 and then averaged their score for each category. For fixed mind-set, scores closer to 1 show that Fellows have more of a growth mind-set regarding STEM intelligence.

For Belongingness, scores closer to 8 indicate Fellows' feeling that they are a member of the STEM community.

The survey was administered at three time points: (1) the beginning of the summer prior to participation in the STAR summer workshops and internships (pre, n = 87), (2) immediately after the summer experience (post, n = 90), and (3) approximately 6 months after the intervention around the closing conference in January (follow up, n = 63). Analysis included comparison of descriptive statistics as well as paired t-tests to compare results from pre to post and post to follow-up. This analysis was used to compare different sets of STAR Fellows who completed surveys at different time points. Due to the multiple comparisons, a more stringent alpha was set to mitigate type 1 errors (significant differences that do not exist but are a result of multiple tests on the same data).

An additional set of questions was added in 2019 that asked STAR Fellows about strategies they may or may not employ for different hypothetical students who may be in their classes. STAR Fellows were asked to rate statements that represented how they would respond to students in a variety of scenarios (e.g., a student who tried hard but still struggled versus a student who did not try hard and did not do well). Strategies included both comfort strategies (praising students, telling them that could do well in other classes) that align with a fixed mind-set as well as effort strategies (telling them they would do better if they put in more effort) that align with a growth mind-set. Each statement was rated on a scale from 1 (Would use most infrequently because it would not meet my instructional goals) to 5 (Would use very frequently because it would meet my instructional goals). Analysis of these questions was conducted with descriptive statistics to look for trends in the relative use of comfort strategies and effort strategies for students at different time points.

Productive Mind-Set Findings

As can be seen in **TABLE 5** below, STAR Fellows' responses to the fixed mind-set measures changed significantly over time. Specifically, participation in STAR led to stronger growth mind-set post-summer compared to pre-summer ($p < .001$) with a small effect size. This difference persisted over time such that the differences between the post-survey and the follow-up survey 5 months later were not significantly different. Thus, we found that despite the passage of time and not actively being in the weekly workshops during the academic year (which served as a mind-set intervention), Fellows persisted in their growth mind-set.

STAR Fellows' responses to the measure of sense of belonging to the STEM community also changed across the summer. Specifically, participation in STAR led to stronger feelings of belonging to the STEM community on the post-survey compared to the pre-survey with a medium effect size. These feelings were not retained over time and diminished significantly with a small effect size from the post-survey to the follow-up survey 5 months later.

TABLE 5

Summary of Mind-Set and Belongingness Scores on Pre- and Post-Survey

Mind-Set Concept	Average Scores and SD on Pre-Survey	Average Scores and SD on Post-Survey	Significant Difference and Effect Size
Fixed Mind-Set ($n = 83$)	1.98 (1.26)	1.55 (1.02)	$p < .001$, $d = .37$
Belongingness ($n = 88$)	5.99 (1.3)	6.86 (1.1)	$p < .001$, $d = .67$

TABLE 6

Summary of Mind-Set and Belongingness Scores on Pre- and Post-Survey

Mind-Set Concept	Average Scores and SD on Post-Survey	Average Scores and SD on Follow-Up Survey	Significant Difference and Effect Size
Fixed Mind-Set ($n = 54$)	1.56 (1.1)	1.59 (1.02)	$p = .891$, $d = .03$
Belongingness ($n = 55$)	6.79 (.97)	6.41 (1.33)	$p < .01$, $d = .4$

During the 2019–2020 academic year, STAR Fellows' responses decreased in comfort strategies (e.g., praising their "students," telling them they had more chances) and increased in effort strategies (e.g., telling them they would do better if they put in more effort) from pre- to post-summer. Fellows' open-ended responses also showed a shift from being afraid to make students feel bad about their performance to more constructive thinking on how they wanted to encourage students to put in more effort. A representative comment from the pre-survey that embedded more of a comfort strategy was, "'I don't think you're putting enough effort toward this class.' I think this could be phrased more gently, in a way that leaves room for understanding that some students may have a valid reason for not putting a lot of effort into a class." A representative comment from the post-survey that showed more effort strategies was, "I would encourage the student to continue to try hard and work on new strategies to help them be successful."

These preliminary results suggest that STAR workshops are effective and enduring interventions to foster growth mind-sets in STEM teachers, and that the effect persists even when teachers are not actively engaged in ongoing professional development around growth mind-set. The results also suggest that STAR is an effective intervention to foster feelings of belonging to the STEM community, but that ongoing support may be necessary to maintain these feelings of belonging once the Fellows leave their research labs. Lastly, the results show that productive mind-sets are reflected in how STAR Fellows plan to react to students and may impact their classroom practice.

Work in Progress: Teacher Retention

In addition to studying teacher effectiveness, an ongoing comparative study has collected state administrative employment records and individual interviews with STAR teachers and comparison non-STAR teachers to determine if STAR Fellows are more likely to be retained in the teaching professions, especially in high-need settings. Researchers used the California Department of Education's classroom census data from the 2012–2013 through the 2017–2018 academic years for credential completers from three pools: STAR alumni, STAR applicants, and credential candidates from the four California State University campuses involved in the first two studies described above, which had high percentages of STAR alumni and applicants. In addition, interviews were conducted with the teachers involved in Study 1 to learn more about retention. Results of this investigation will be reported elsewhere.

Discussion and Conclusion

Discussion and Reflection

The STAR Program is the largest program in the United States that provides mentored research experiences with pedagogical support for prospective and preservice teachers. With an aim to prepare students to enter the STEM teacher workforce, it is a unique undergraduate research experience, distinct from the intention of other undergraduate research programs that work to prepare students to enter the STEM workforce. Also, while traditional TREs often focus on supporting inservice teachers in growing in their knowledge and skills necessary to engage their students in authentic and appropriate research practices, the STAR Program exists at the intersection of undergraduate research and teacher professional development. STAR serves a group of self-selected students who may have limited research experience, do not have the intention to become professional researchers or work in STEM-related industry, and have not yet experienced classrooms as independent teachers. Through these studies, we provide evidence that engagement of preservice and early-career teachers in paid summer research with intentional professional development can lead to positive impacts on teachers' classroom practices and growth mind-sets. Specifically, teachers who participated in STAR may prepare students for STEM careers differently and support enhanced perseverance in STEM. There is evidence from teachers' narratives that these practices are directly related to their personal experiences in the STAR Program.

STAR Fellows deepened their growth mind-set and sense of belonging in STEM and changed their views on how to best support students to encourage more effort, which may help students persist when faced with challenges. The study of student achievement revealed no significant gains in student achievement that were measured in mathematics and science state tests (both by district and in aggregate), while students of STAR science teachers in a single district performed less well in English language arts than did their matched peers. While trying to replicate other studies of student achievement, this study employed a more general high-stakes assessment that was not aligned to a specific content area. This is an area that would benefit from additional research focused on creating more closely aligned assessments that can detect student achievement on relevant skills and topics consistent with research practices. Overall, these findings not only support other work demonstrating positive outcomes of TREs (e.g., Krim et al., 2019; Miranda & Damico, 2013; Southerland et al., 2016) but also add to the few studies that include student outcomes as a measure of TRE impact.

We wish to highlight limitations and nuances of these studies to frame not only our findings but also challenges that researchers may face in doing related research. The STAR Program provides research experiences as a component of teacher preparation and induction (participants must not have taught full time when they first participate in the program and cannot be beyond their second year of full-time teaching). Thus, these studies, and any others, involving preservice participants lack the ability to collect baseline data of teacher effectiveness prior to participation in the intervention. With inservice TRE programs, surveys, interviews, classroom observations, instructional artifact collection, and student achievement scores for a specific teacher can all be collected in advance of participation in the research experience and then compared with changes following the intervention. While we strongly encourage further research studies of inservice teachers that make use of baseline data, our study of preservice teachers utilized propensity matching with prospective comparison teachers instead of pre- and post-participation comparisons to investigate teacher effectiveness. Another nuance is the limited number of STAR teachers included in Study 1 and Study 2. This was driven primarily by resources, capacity, and limited tracking of STAR alumni teacher placements prior to this investigation. While the project worked with the four CSU campuses that have involved the largest numbers of STAR participants, the STAR alumni are in actuality teaching in positions across the state and the country. Of all teachers in the five districts that partnered is this work, we were only able to identify and obtain consent from eight STAR alumni. With teacher employment records from 2010 to 2018 obtained through our teacher retention efforts, we are now in a better position to identify districts with higher numbers of STAR alumni, but this information was not available in summer 2018 when the case study on teacher effectiveness was being implemented.

In the second study, only five of our eight district partners were able to share the requested student achievement data due to district resources. Additionally, in cases involving STAR alumni who primarily taught senior or junior/senior classes, junior year student achievement data were not relevant or useful. Overall, we want to highlight the complexity of gathering this targeted teacher and student data related to a research program for teachers. Early planning and strong partnerships were required for the work we were able to achieve, and we encourage other researchers to ensure sufficient resources are available for the gathering of such data. Although the findings about mind-set and belongingness within the same year as the STAR experience are promising, they do not include impacts on teachers' classrooms. This is another avenue of research that will be important to investigate alongside those of other research classroom practices. We plan to study more classroom outcomes related to the infusion of productive mind-set in our summer education workshops to better understand the intersection between belongingness in STEM, productive mind-set, and student outcomes.

Taken together, these three studies demonstrate the value of TREs for preservice teachers. Literature already demonstrates the value of TREs for inservice teachers, both for classroom practice and student outcomes, and this work contributes the finding that research experiences early in a teaching career can also have a measurable impact on teaching in the field. All three studies will be of interest for programs that support, or aim to support, TREs for preservice teachers as well as for researchers who want to undertake similar work. As next steps for this kind of research, we plan to explore new ways to investigate student achievement, potentially through custom assessments that are better aligned to practices. Additionally, classroom observations will add more insights into teachers' practices and students' work. The innovations of summer workshops to augment teacher research experiences to support teachers' integration of their research experiences through the lens of productive mind-set provides several avenues for future work.

Implications for the Field and Future Work

The first two studies described here represent some of the initial efforts by a preservice TRE program to extend assessment beyond teacher self-report to include data from students in order to provide evidence of teacher effectiveness and classroom practices. While the field of inservice TRE programs has a few comparative studies that go beyond self-report data analysis (e.g., Silverstein et al., 2009), the field of preservice TRE programs is even more limited. While the case study sample for Study 1 is small, the study demonstrates that student surveys can be used along with triangulation with teacher interviews to meaningfully discriminate between teachers in terms of student perceptions of classroom learning constructs. Indeed, in another study of STAR, a broad sample of 136 STAR alumni and 60 STAR applicants responded to a teacher survey involving many of these same constructs, and we were not able to differentiate between the two groups from self-report data. Thus, we recommend the strategy of collecting student perceptions to further inform perspectives on the classroom experience integrated over the course of the year.

In teacher interviews conducted during the case studies, amongst both STAR and non-STAR teachers, we found that most of the teachers had some experience in industry or research, leading us to new questions about these experiences and the role, or lack of role, of supporting professional development. Because nominal research experience across a spectrum of opportunities exists in STEM education, more attention to the role that an emphasis on metacognition and professional development plays in teacher preparation is warranted.

Teacher interview data was helpful for discerning some preliminary qualitative factors behind measurable constructs for teacher effectiveness found in student survey findings. These included STEM career awareness and perseverance, as well as the ability to integrate research into the classroom. Surveys or focus groups with students who completed surveys for the project may have been additionally useful. Each of these measures represents meaningful avenues for any future research to examine the impact of research experiences for teachers. Future research could also be conducted involving analysis of student artifacts and classroom observations to better understand learning experiences related to teachers' classroom practices.

The success of this project to secure student achievement data for nine STAR teachers and comparison teachers in five districts for teachers from 2010 to 2018 represents a significant accomplishment. Additionally, the data we are currently collecting on employment and retention represents a significant partnership with the California Department of Education. To successfully do this type of work, we recommend researchers work with state departments of education to learn the extent to which access can be obtained to external, third-party employment data.

In addition to the results described above, ongoing evaluation findings have demonstrated that participants in the STAR Program develop and strengthen an identity as teacher-researchers, which is a central goal of the program. This shift in participants' identity from either a STEM teacher or a STEM researcher toward that of a teacher-researcher is a significant outcome of STAR. These preliminary findings align with NSF's own recent strategy for the Robert Noyce Scholarship Program to encourage grantees to provide research experiences for teachers (Marrongelle, 2021). This parallels other work in identity formation in both science and education and contributes to efforts to better understand how experiences impact identity. We also suggest that investigations of the development of STEM identity amongst the students of teacher-researchers may be a productive avenue of research.

Our findings are exciting, and hopefully aspirational, for those who actively research TREs and work with Noyce Fellows. Additionally, this work is valuable to STEM teacher educators, as we provide insight into new ways to prepare teachers to engage students in research practices, foster productive struggle and mind-set, and show evidence of the potential efficacy of this model of teacher professional development and preparation.

References

Anderman, E. M., Eccles, J. S., Yoon, K. S., Roeser, R., Wigfield, A., & Blumenfeld, P. (2001). Learning to value mathematics and reading: Relations to mastery and performance-oriented instructional practices. *Contemporary Educational Psychology, 26*(1), 76-95.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2), 113-125.

Autenrieth, R., Lewis, C., & Butler-Purry, K. (2018). Enrichment experiences in engineering (E3) summer teacher program: Analysis of student surveys regarding engineering awareness. *Journal of STEM Education: Innovations and Research, 19*(4).

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107*(5), 1860-1863.

Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances, 5*(2).

Dweck, C. S. (2013). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.

Dweck, C. (2016). What having a "growth mindset" actually means. *Harvard Business Review, 13,* 213-226.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273. https://psycnet.apa.org/doi/10.1037/0033-295X.95.2.256

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science, 14*(3), 481-496.

Engeström, Y. (2015). *Learning by expanding.* Cambridge University Press.

Evans, G., & Crippen, K. (2020). MRET badges: Using micro certifications to scaffold an engineering laboratory research experience for elementary teachers. *ASTE International Conference* (pp. 1-6).

Feldman, A., Divoll, K. A., & Rogan-Klyve, A. (2013). Becoming researchers: The participation of undergraduate and graduate students in scientific research groups. *Science Education, 97*(2), 218-243.

Feldman, A., & Özalp, D. (2019). Science teachers' ability to self-calibrate and the trustworthiness of their self-reporting. *Journal of Science Teacher Education, 30*(3), 280-299.

Friday Institute for Educational Innovation (2012). *Teacher Efficacy and Attitudes Toward STEM Survey.* Author.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*(6), 645-662.

Good, C., Mangels, J., & Evelo, A. (2015, February). *Contingencies of belonging protect against the effects of stereotype threat on learning.* [Paper presentation]. Society for Personality and Social Psychology 16th Annual Meeting, Long Beach, CA.

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology, 102*(4), 700-717. https://doi.org/10.1037/a0026659

Good, C., & Regnier, M. (2016). 7 things growth mindset is not. https://medium.com/turnaround-for-children/7-things-growth-mindset-is-not-2f3539660f84

Gooyabadi, M. & Good, C. (2016, February). It's your fault, not mine: The hidden costs of growth mindset. Presentation at the Society for Personality and Social Psychology, San Diego, CA.

Grant, H., & Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology, 85*(3), 541-553.

Hayes, K. N., Lee, C. S., DiStefano, R., O'Connor, D., & Seitz, J. C. (2016). Measuring science instructional practice: A survey tool for the age of NGSS. *Journal of Science Teacher Education, 27*(2), 137-164.

Keller, J. M., Buxner, S. R., Donnelly-Hermosillo, D., Bailey, E., Citkowicz, M., Horvath, L., Moreno, D., Yisak, M., Zhu, B., Fulbeck, E., Sessoms, D., Vokos, S., Chen, C., & Pardo, M. (2022). Student outcomes of practicing teachers who engaged in a pre-service or early career teacher research experience program. Manuscript submitted for review.

Krim, J. S., Coté, L. E., Schwartz, R. S., Stone, E. M., Cleeves, J. J., Barry, K. J., Burgess, W., Buxner, S.R., Gerton, J.M., Horvath, L., Keller, J.M., Lee, S. C., Locke, S.M., & Rebar, B. M. (2019). Models and impacts of science research experiences: A review of the literature of CUREs, UREs, and TREs. *CBE—Life Sciences Education, 18*(4).

Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science, 347*(6219), 262-265.

Luera, G., & Murray, K. (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*(4), 303-331. https://doi.org/10.5408/14-064.1

MacFadden, B. J., Pirlo, J., Abramowitz, B., Killingsworth, S., & Ziegler, M. (2020). *Authentic field experiences for STEM teachers: Collecting Florida fossils*. Cold Spring Harbor Laboratory. http://dx.doi.org/10.1101/2020.07.02.184077

Mangels, J. A., Butterfield, B., Lamb, J., Good, C., & Dweck, C. S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social Cognitive and Affective Neuroscience*, 1(2), 75-86. https://doi.org/10.1093/scan/nsl013

Marrongelle, K. (2021). Dear Colleague Letter: Noyce research experiences in science, technology, engineering, and mathematics (STEM) settings for pre-service teachers. https://www.nsf.gov/pubs/2021/nsf21086/nsf21086.jsp

Miranda, R. J., & Damico, J. B. (2013). Science teachers' beliefs about the influence of their summer research experiences on their pedagogical practices. *Journal of Science Teacher Education, 24*(8), 1241-1261.

Ragusa, G., & Juarez, C. (2017, June). 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. Paper presented at: 2017 ASEE Annual Conference & Exposition (Columbus, OH). Retrieved from https://peer.asee.org/27710

Rattan, A., Good, C., & Dweck, C. S. (2012). "It's ok - Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology, 48*(3), 731-737.

Reynolds, D., Yazdani, N., & Manzur, T. (2013). STEM high school teaching enhancement through collaborative engineering research on extreme winds. *Journal of STEM Education, 14*(1), 12-19.

Shanahan, M.-C., & Bechtel, R. (2020). "We're taking their brilliant minds": Science teacher expertize, meta-discourse, and the challenges of teacher–scientist collaboration. *Science Education, 104*(2), 354-387. https://doi.org/10.1002/sce.21550

Silverstein, S. C., Dubner, J., Miller, J., Glied, S., and Loike, J. D. (2009). Teachers' participation in research programs improves their students' achievement in science. *Science, 326*, 440. http://doi.org/10.1126/science.1177344

Smarter Balanced Assessment System, https://www.cde.ca.gov/ta/tg/sa/

Smith, T., Brumskill, R., Johnson, A., & Zimmer, T. (2018). The impact of teacher language on students' mindsets and statistics performance. *Social Psychology of Education, 21*(4), 775-786.

Southerland, S. A., Granger, E. M., Hughes, R., Enderle, P., Ke, F., Roseler, K., Saka, Y., & Tekkumru-Kisa, M. (2016). Essential aspects of science teacher professional development. *AERA Open, 2*(4), 233285841667420. https://doi.org/10.1177/2332858416674200

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613.

TIMSS 2007 Assessment Frameworks. Copyright © 2007 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from 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.

Wade-Jaimes, K., Cohen, J. D., & Calandra, B. (2019). Mapping the evolution of an after-school STEM club for African American girls using activity theory. *Cultural Studies of Science Education, 14*(4), 981-1010.

Westerlund, J. F., García, D. M., Koke, J. R., Taylor, T. A., & Mason, D. S. (2002). Summer scientific research for teachers: The experience and its effect. *Journal of Science Teacher Education, 13*(1), 63-83.

Williams, D. A., & King, P. (1980, December 15). Do males have a math gene? *Newsweek, 73*. Retrieved from http://beck2.med.harvard.edu/week6/Newsweek Time.pdf

Yang, Y., Liu, X., & Gardella Jr., J. A. (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*(7), 1028-1057.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., . . . & Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature, 573*(7774), 364-369.

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist, 47*(4), 302-314.