

# Models and Impacts of Science Research Experiences: A Review of the Literature of CUREs, UREs, and TREs

Jessica S. Krim,<sup>†\*</sup> Laleh E. Coté,<sup>‡§</sup> Renée S. Schwartz,<sup>||</sup> Elisa M. Stone,<sup>¶</sup>  
 Jessica J. Cleaves,<sup>#</sup> Kelly J. Barry,<sup>@</sup> Wilella Burgess,<sup>\*\*</sup> Santyn R. Buxner,<sup>††</sup>  
 Jordan M. Gerton,<sup>‡‡‡</sup> Lawrence Horvath,<sup>§§</sup> John M. Keller,<sup>||||</sup> Soon Chun Lee,<sup>¶¶</sup>  
 Sharon M. Locke,<sup>###</sup> and Bryan M. Rebar<sup>@@</sup>

<sup>†</sup>Curriculum and Instruction, <sup>@</sup>Biological Sciences, and <sup>##</sup>Environmental Sciences and STEM Center for Research, Education, and Outreach, Southern Illinois University, Edwardsville, Edwardsville, IL 62026; <sup>‡</sup>Workforce Development & Education, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; <sup>§</sup>Graduate Group in Science & Mathematics Education and <sup>¶</sup>CalTeach, University of California, Berkeley, Berkeley, CA 94720; <sup>||</sup>Science Education, Georgia State University, Atlanta, GA 30306; <sup>#</sup>Center for Science and Mathematics Education and <sup>††</sup>Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112; <sup>\*\*</sup>Evaluation and Research Learning Center, Purdue University, West Lafayette, IN 47907; <sup>†††</sup>Teaching/Learning and Sociocultural Studies, University of Arizona, Tucson, AZ 85721; <sup>§§</sup>Secondary Education and Center for Science and Math Education, San Francisco State University, San Francisco, CA 94132; <sup>||||</sup>Fiske Planetarium, Astrophysical and Planetary Sciences, University of Colorado, Boulder, Boulder, CO 80309; <sup>¶¶</sup>STEM Education, Wichita State University, Wichita, KS 67260; <sup>@@</sup>STEM CORE, Science, Technology, Engineering, and Math Careers through Outreach, Research, & Education, University of Oregon, Eugene, OR 97403

## ABSTRACT

In efforts to increase scientific literacy and enhance the preparation of learners to pursue careers in science, there are growing opportunities for students and teachers to engage in scientific research experiences, including course-based undergraduate research experiences (CUREs), undergraduate research experiences (UREs), and teacher research experiences (TREs). Prior literature reviews detail a variety of models, benefits, and challenges and call for the continued examination of program elements and associated impacts. This paper reports a comprehensive review of 307 papers published between 2007 and 2017 that include CURE, URE, and TRE programs, with a special focus on research experiences for K–12 teachers. A research-supported conceptual model of science research experiences was used to develop a coding scheme, including participant demographics, theoretical frameworks, methodology, and reported outcomes. We summarize recent reports on program impacts and identify gaps or misalignments between goals and measured outcomes. The field of biology was the predominant scientific disciplinary focus. Findings suggest a lack of studies explicitly targeting 1) participation and outcomes related to learners from underrepresented populations, 2) a theoretical framework that guides program design and analysis, and, for TREs, 3) methods for translation of research experiences into K–12 instructional practices, and 4) measurement of impact on K–12 instructional practices.

## INTRODUCTION

As today's world continues to be shaped by science and technology, there is a pressing need to improve public understanding of what constitutes “science” and scientific practices (National Research Council [NRC], 2012). A report commissioned by the NRC (Duschl *et al.*, 2007) emphasized the need to improve science, technology, engineering, and mathematics (STEM) literacy among all citizens while also encouraging underrepresented groups to aspire to and succeed in STEM fields. Teachers are called to engage students in investigative practices modeled after those of scientists,

James Hewlett, *Monitoring Editor*

Submitted Apr 4, 2019; Revised Aug 5, 2019;  
 Accepted Sep 11, 2019

CBE Life Sci Educ December 1, 2019 18:ar65  
 DOI:10.1187/cbe.19-03-0069

\*Address correspondence to: Jessica S. Krim  
 (jkrim@siue.edu).

© 2019 J. S. Krim *et al.* CBE—Life Sciences Education © 2019 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

engineers, and mathematicians; to enhance students' conceptions of nature of science and the nature of scientific, engineering, and mathematics practices; and to foster STEM identities to educate learners of all ages toward these proficiencies. *A Framework for K–12 Science Education* (NRC, 2012) and the Next Generation Science Standards (NGSS Lead States, 2013), encourage educators in the United States to make science learning experiences more authentic by challenging their students to creatively solve real-world problems and address current scientific issues.

Learning experiences need to reflect the relevance, curiosity, and inspiration of the learners themselves to attract them to STEM fields and retain them through degree completion. For science, much of this work has focused on “real” or “authentic” science experiences, by which educators often mean science that reflects practices of scientists and engineers and engagement in scientific habits of mind (Roth, 1995). To meet the challenge of helping students “learn science in a way that reflects how science actually works” (NRC, 1996, p. 214), many countries around the world are recommending learners be engaged in designing and conducting scientific investigations (Hasson and Yarden, 2012; Crawford, 2014). Among the most recent recommendations from the United States, the NGSS emphasize a three-dimensional approach to K–12 science education focused on core disciplinary ideas, scientific practices, and cross-cutting concepts.

Despite decades of recommendations to involve learners in scientific activities that model authentic science, K–12 teachers still struggle to integrate these practices across their curricula (Capps and Crawford, 2013; Crawford, 2014). One commonly stated reason for teachers' continued challenges is they typically lack firsthand scientific research experience in an authentic setting, either in laboratory classes or within academic or industry research settings (Schwartz and Crawford, 2004; Sadler *et al.*, 2010). In response, multiple programs have engaged teachers and future teachers in scientific research experiences (Sadler *et al.*, 2010). For many years the National Science Foundation (NSF) has funded Research Experiences for Teachers (RET) programs to engage practicing teachers in authentic science experiences that will translate into enhanced learning experiences for K–12 students. In parallel, a growing number of programs involve undergraduate students in scientific research projects in a laboratory or industrial setting—undergraduate research experiences (UREs; Linn *et al.*, 2015; National Academies of Sciences, Engineering, and Medicine [NASEM], 2017), also referred to as research experiences for undergraduates (National Science Foundation, 2019), research apprenticeships (Sadler *et al.*, 2010), or research internships (Auchincloss *et al.*, 2014)—or in higher education classrooms—course-based undergraduate research experiences (CUREs; Corwin *et al.*, 2015). Because a number of undergraduate students in UREs and CUREs may become science teachers after additional teacher preparation and credentialing, either as undergraduates or in postbaccalaureate programs, what they learn in these research experiences is critical for their future careers as teachers. Also, a variety of programs place teachers in industry jobs over the summer with the intention of providing expanded opportunities designed to increase teachers' abilities to bring workplace-related 21st-century skills to their classrooms

(Ignited, 2018; University of Arizona, 2018; Virginia Science and Technology Campus, 2018).

## CALLS TO ACTION

Calls to action for more intentional program and course design, as well as systematic research targeting CUREs, UREs, and teacher research experiences (TREs, including NSF RET programs and other teacher programs funded by other sources), come from several comprehensive reviews of published studies in these areas (Sadler *et al.*, 2010; Corwin *et al.*, 2015; Linn *et al.*, 2015) and reports (Auchincloss *et al.*, 2014; NASEM, 2017). Sadler *et al.*'s (2010) review reported on 53 empirical papers published between 1961 and 2008 that focused on science research experiences for secondary students, undergraduates, and teachers, both preservice and in-service. Sadler and colleagues examined learning outcomes associated with participation in these research activities, including career aspirations, nature of science, scientific content knowledge, self-efficacy, intellectual development, skills, satisfaction, discourse practices, collaboration, and changes in teacher practices. Moreover, Sadler and McKinney (2010) highlighted studies specifically involving undergraduate students and reported similar gains in career goals, confidence, content knowledge, and skills. Both of these reviews reported that science apprenticeships or research programs vary substantially regarding the length of experience, student epistemic engagement, and embedded supports for learning. They recommended that further studies use valid and reliable instruments to explore relationships among program components and outcomes and to identify and design experiences that provide epistemic challenges to participants in order to maximize potential gains.

The Corwin *et al.* (2015) review examined 14 studies on CUREs and 25 studies on UREs published between 2006 and 2014. The authors determined the level of empirical support within each study and the alignment of activities and outcomes in order to create minimodels based on common outcome themes. They then developed a more extensive, comprehensive model containing “hubs” that provided a framework for better understanding the research. Corwin *et al.* (2015) emphasized the importance of developing a program model before selecting which outcomes to measure and using a three-phase evaluation system in which a variety of outcomes may be measured at appropriate time points.

In their 2015 review, Linn and colleagues examined 60 empirical studies published between 2010 and 2015 that described UREs (Linn *et al.*, 2015). They found that most of the studies relied on self-reported gains gathered through surveys or interviews. Only four studies used additional instruments to measure gains in undergraduates' research capabilities or conceptual understanding. The authors called for rigorous research methods to identify ways to design UREs that promote integrated understanding, specifically those with systematic, iterative studies that build in multiple indicators of success.

A recent NASEM report (2017) outlines a set of recommendations that reiterate the need for well-designed studies that have the capacity to surface causal relationships among components of UREs and for evidence-based URE design and refinement. The report provides suggestions for improved data collection; professional development opportunities for faculty; and mentor training for graduate students, postdoctoral

researchers, and faculty who work with undergraduate researchers. The report also calls for strengthening collaborative partnerships within and across institutions to sustain URE efforts.

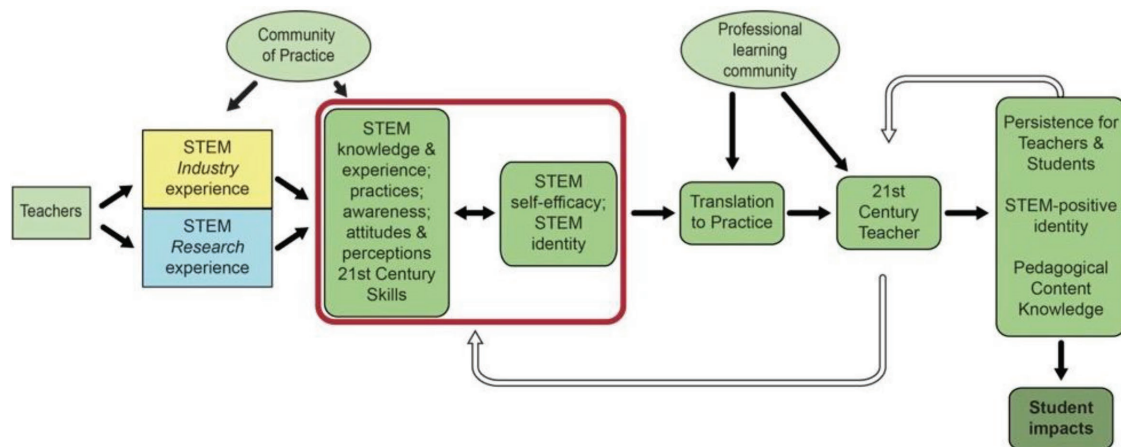
Moreover, policy makers have tasked educators with repairing and broadening the pipeline into STEM careers, including opening accessibility to the STEM pipeline across underrepresented groups. Engaging all students in authentic science is critical due to the widely acknowledged achievement gap between students of low and high socioeconomic status (Astin and Oseguera, 2004; Hoxby and Avery, 2013) and the underrepresentation of women, persons of color, and persons with disabilities in science and engineering (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, 2011). Linn *et al.* (2015) identify elements to strengthen mentoring and ensure that research experiences meet the needs of diverse students, including socialization, emotional support, cultural norming, remedying gaps in preparation, and promoting science identity. NASEM (2017) summarizes literature that supports a correlation between the quality and frequency of mentee–mentorship interactions with positive undergraduate student outcomes, including for underrepresented students, and calls for additional research in this area. Sadler *et al.* (2010) call attention to four studies in particular (Nagda *et al.*, 1998; Davis, 1999; Campbell, 2002; Ponjuan *et al.*, 2007) that highlight the need for program designers and evaluators to give explicit attention to program features and experiences that target women and students of color.

Research also shows that learners' science content knowledge, skills, beliefs, and attitudes toward science, including interests in careers in science fields, can be enhanced through engaging in science research experiences (Sadler *et al.*, 2010). While there are indications that teachers' participation in

research impacts K–12 student learning (Silverstein *et al.*, 2009), the research demonstrates that teacher or future teacher participants generally require more than a research experience, no matter how long, to impact instructional practice (Schwartz and Crawford, 2004; Schwartz *et al.*, 2004, 2010; Sadler *et al.*, 2010; Enderle *et al.*, 2014). TRE programs that purposefully guide teachers to consider connections between authentic science contexts and their science teaching and provide support for the translation of science research experiences into lesson development are more likely to impact teaching practices in ways that enhance student engagement in scientific practices (Sadler *et al.*, 2010). Together, these calls to action focus on clarifying impactful program elements through valid measures and identifying models that are effective for a variety of participants in order to increase interest and participation in STEM.

### THE COLLABORATIVE AROUND RESEARCH EXPERIENCES FOR TEACHERS

The work presented here is the outcome of efforts coordinated by the Collaborative Around Research Experiences for Teachers (CARET). This collaborative began in January 2015 and has subsequently grown to include representatives from 11 institutions of higher education, two nonprofit institutions, and one national laboratory, who aim to better understand the impacts of research and industry experiences on teacher effectiveness and retention. Toward this end, CARET developed a preliminary model for TREs involving both preservice and in-service teachers (Figure 1; discussed below). The group also recognized that many future teachers engage in research opportunities through UREs and CUREs. Consequently, this review includes publications with respect to CUREs, UREs, and TREs, with particular emphasis on teacher development. In addition



**FIGURE 1.** The CARET model describes elements of STEM TREs and their intended outcomes. Many STEM teachers engage in STEM research experiences, in either academic or industry settings. They do so as in-service teachers (primarily during the summer when they are not teaching in the classroom) and as preservice teachers (primarily as undergraduates before becoming classroom teachers). Undergraduate and teacher researchers are guided to engage in research with experienced researchers in a Community of Practice (e.g., faculty, postdoctoral researchers, and graduate students in a laboratory setting). Outcomes of these experiences include increased STEM knowledge and experience, scientific research practices, career awareness, and STEM self-efficacy and identity. Programs designed for TREs typically are guided to translate their research into classroom practices that include curricular development by a Professional Learning Community, which leads to improvements in STEM teaching and learning, represented here as a “21st Century Teacher,” and include outcomes such as increased persistence in STEM teaching and pedagogical content knowledge. Solid arrows indicate a direct connection; red outline includes direct outcomes from research experiences; open arrows indicate feedback loops; and light green shading represents the individuals or groups involved (green shading deepens for initial and longer-term outcomes).

to this literature review, the collaborative has created and is currently piloting a shared metric instrument for use across multiple TRE programs (data not included in this review).

## PURPOSE OF THE REVIEW

The purpose of this comprehensive literature review is to gain an understanding of reported CURE, URE, and TRE program features, targets, and outcomes. We examined relevant papers published from 2007 to 2017 to capture a “state of the field” with respect to types of programs, participants, program elements, assessment measures, outcomes, and theoretical frameworks. In particular, this review targets studies of TRE programs published since the review by Sadler *et al.* in 2010. Moreover, attention is given to CURE, URE, and TRE studies published between 2014 and 2017 to enable comparison across the most recent literature for these three program types.

The research questions that guided this literature review are 1) How has the research literature about science research experiences for teachers and undergraduates changed over the past decade? 2) What is the involvement of, and impact on, participants from underrepresented groups in science research experiences? 3) How are science research experiences impacting teachers and science teaching? 4) How are studies of research experience programs being conducted and what type of data sources are in use? From these results, we identify trends and gaps and make recommendations for future research.

## THEORETICAL FRAMEWORK

Situated learning and communities of practice serve as the theoretical frameworks guiding this review (Lave and Wenger, 1991; Wenger, 1998). A “community of practice” framework (Crawford, 2014), in which learners can engage in a variety of social and scientific activities (e.g., questioning, investigating, communicating, critiquing), is used to describe the laboratory or industry community in which undergraduate or teacher researchers develop an understanding of their scientific research practices. Situating participants in a scientific community of practice provides an opportunity for learning in an authentic context and is essential for transitioning from the periphery to community membership (Lave and Wenger, 1991).

Considering these frameworks, and the results of the prior literature reviews, we developed the CARET model. The CARET model posits that teachers or aspiring (preservice) teachers who engage in STEM research or STEM industry experience will demonstrate shifts in professional and pedagogical practices and identity/self-efficacy (Figure 1). The CARET model was developed by the collaborative research team to provide guidance in the literature search and review. We do not assert that the model is inclusive of all impactful features or strategies associated with research or industry experiences for teachers; nor is this study a systematic test of the model. For our purposes, the model serves as a starting point to articulate, challenge, and refine our understanding of CUREs, UREs, and TREs in order to examine the literature.

While most of the past literature and reviews cited here focus on science research experiences, and in particular, research experiences in the field of biology, we chose to use the broader concept of “STEM” in the CARET model to be inclusive of the scientific, technological, engineering, and mathematical disciplinary focus areas for both the experiences as well as the

variety of subject areas taught by K–12 teacher participants. By doing so, we acknowledge that research experiences can include multidisciplinary and integrated experiences that are important for professional teacher learning (Roehrig *et al.*, 2012).

Critical components of the CARET model are the STEM research or STEM industry experience itself and a “translation to practice” experience. Within the model, common characteristics must be present in a STEM research or industry experience, like exposure to scientific practices, collaboration, iteration, discovery, and relevant research. STEM research experiences can take place in any setting in which scientists and teachers conduct original STEM research, and likewise, STEM industry experiences in settings in which STEM skills are applied to advance the interests or goals of the industry.

The CARET model identifies a community of practice (Lave and Wenger, 1991), that is, a group of people involved in the research and who are participants in authentic scientific research, such as mentors who are faculty or senior scientists, postdoctoral researchers, graduate students, as an essential component of the research experience. This model posits that a community of practice is vital in achieving immediate STEM-related outcomes for teachers, such as increased understanding of STEM knowledge and practices, which can lead to enhanced STEM self-efficacy and STEM-positive identity. Note that goals and intended outcomes of the STEM research or industry experience for teachers are parallel to those for undergraduate researchers (e.g., see “Goals for students participating in UREs,” NASEM, 2017, p. 71).

Unique to the TRE experience, teacher researchers must also have the opportunity to translate their knowledge and skills into their classrooms (Schwartz and Crawford, 2004; Sampson *et al.*, 2011). This experience can be a formal process that involves actions such as structured reflections, group discussions, and supported lesson development. The experience may also be less formal, or on an individual basis, such as teachers being prompted to consider how they might translate this knowledge and skills into their classrooms. The model also acknowledges the role that professional learning communities (PLCs) can play in supporting and facilitating instructional change (Dogan *et al.*, 2016) by way of a variety of activities, such as collaborative curricular development or networking and communicating among peers who are participating in the experience together.

The idea of a “21st-century teacher” is named in the spirit of the Framework for 21st Century Learning (Partnership for 21st Century Learning, 2018) and the Next Generation Science Standards (NGSS Lead States, 2013; NRC, 2013). This term implies that the teacher, after the STEM research or industry experience, may be better equipped to teach in a manner that not only facilitates student learning outcomes in his or her key STEM discipline but also in such areas as critical thinking, evaluating perspectives, and integrating the use of supportive technologies.

As a result of the research experience and support to translate the experience into instructional practice, teachers become part of a PLC (Dogan *et al.*, 2016) that enables them to more effectively integrate science subject matter and science and engineering practices within their localized contexts. The CARET model includes PLCs as a transformative factor needed for developing pedagogical content knowledge and other teacher outcomes like



self-efficacy and positive STEM identity. The CARET model also highlights the long-term goal of research experiences for teachers: outcomes associated with K–12 students, such as STEM learning, identity, and persistence. The model illustrates feedback loops that point to the iterative nature of teacher growth and experience as continuous cycles, as well as two key sociocultural agents, communities of practice and PLCs.

## METHODOLOGY

### Inclusion and Exclusion Process

A systematic search for papers relevant to research experiences was undertaken using the following key words in various combinations: “science,” “STEM,” “math,” “biology,” “chemistry,” “physical sciences,” “geology,” “geosciences,” “physics,” “earth sciences,” “engineering,” “energy,” “computer science,” “medical,” “materials sciences,” “course-based,” “undergraduate(s),” “teacher(s),” “research,” “classroom-based,” “experience(s),” and “internship(s).” Search engines used were ERIC, ERIC-EBSCOhost, ERIC Proquest, Google Scholar, and Web of Science. An organizational spreadsheet was created that outlined a search hierarchy, guided the overall search process, and avoided redundancy. This search hierarchy specified the order in which search engine and key word combinations productively targeted relevant papers. Examples of productive combinations of key words in different search engines included “chemistry undergraduate research experiences” (Web of Science), “science undergraduate research experiences” (Web of Science), STEM “undergraduate research internships” (Google Scholar), “research experiences for teachers” (ERIC-EBSCO).

To narrow the field, the search included only papers published since the most recent reviews; for TREs, search terms containing the key word “teacher” included papers between 2007 and 2017 (Sadler *et al.*, 2010), and for CUREs and UREs, search terms containing the key word “undergraduate” included papers between 2014 and 2017 (e.g., Auchincloss *et al.*, 2014; Corwin *et al.*, 2015; Linn *et al.*, 2015). The initial database search yielded more than 600 unique papers based solely on the titles of the papers and were further screened based on the abstracts. This preliminary collection of 600 papers was refined to 450 and then 307 based on inclusion criteria of 1) peer reviewed; 2) related to CURE, URE or TRE programs; and 3) focused on STEM disciplinary research. The most common reasons for exclusion at this stage included the publication not being related to a CURE, URE or TRE program; the full paper was not accessible (steps were taken to obtain the full text whenever possible); the data collection or analysis had not yet taken place; the publication was a full book (and coding was beyond our capacity). “Data cleaning” was achieved by several team members examining any categories (e.g., item 7: study methods) with blank or unclear entries and filling these in whenever possible by revisiting the relevant paper. Further refinements were based on year of publication to allow fair comparison (2014–2017;  $n = 268$ ) and study type (data driven;  $n = 177$ ). Details of these additional refinements are provided in the following sections.

### Analysis

At least one of 13 members of the CARET coding team (“coders”) coded each of the 450 papers. An initial coding scheme was developed by building on categories used by Linn *et al.* (2015) on undergraduate research. A number of new categories and

codes were added to account for the broadened scope of this literature review, which includes TREs. The group refined the coding guide to align with the CARET model (Figure 1) by introducing several additional codes to account for emergent themes we deemed relevant but the existing scheme had not yet captured.

Using a group calibration process, wherein all members of the coding team coded a single paper and discussed resulting codes over conference calls to reach consensus, allowed for further refinement in defining each coding category. During this process, the codebook was refined to reflect emergent themes as well as collapsed themes due to commonalities. This process was conducted sequentially with a new paper every 2 weeks over a 6-month period until coders consistently reached consensus. The resulting coding guide has 24 items. It is noteworthy that the coding team comprised 13 geographically dispersed members (all located in the United States) who represented a variety of science-related fields. Challenges to establishing intercoder agreement were similar to those reported by others (Kowalski *et al.*, 2018). The process of reducing and clarifying the coding guide as a team was essential for assuring intercoder agreement.

Coders received papers in batches of 10. Helpful coding tools included a detailed coding dictionary that thoroughly defined each coding term and the agreed-upon criteria for placing a paper into one or more particular categories, as well as an electronic coding form (Google Form) that collected coding data into a single spreadsheet. The Google Form permitted coders to view later and make edits to their data as deemed necessary based on group discussions. These procedures facilitated a systematic coding process by multiple coders and gave the group the opportunity to review the data together in an electronic format. As a continued check for intercoder reliability, 16 papers were independently coded by two to five coders, and results were compared. Any discrepancies were addressed by the whole group to reach consensus. See the Supplemental Material for coding guide/dictionary and Google Form survey.

After the coding guide was finalized, the number of relevant papers was reduced to 307 for full analysis. After the 307 papers were coded, descriptive statistics allowed basic comparison by percentages and frequencies, cross-tabulation, and multiple-response using SPSS v. 21. We conducted comparisons within and across type of program (CURE, URE, TRE). A Pearson’s chi-squared test was used for the statistical tests reported.

## Results

We present notable findings across program types and study type, including participant number, identification of participants from underrepresented groups, duration and intensity of research experience, study methodology, outcomes, and author-identified theoretical framework. Additionally, we present more specific results for data-driven studies.

### Program and Study Type

In the final sample of 307 papers, 113 focused on CUREs (published 2014–2017), 85 on UREs (published 2014–2017), and 61 on TRE programs (published 2007–2017); 21 focused on a combination of more than one program type; and 27 were considered “other” due to an inability to determine a clear category.

To focus our analysis on recent literature within the same time period and facilitate comparisons across *program* types, we removed the 39 TRE papers published from 2007–2013

**TABLE 1. Number of participants across all data-driven studies<sup>a</sup>**

No. of participants	TRE <i>n</i> = 22	CURE <i>n</i> = 72	URE <i>n</i> = 65	Combination <i>n</i> = 9	Other <i>n</i> = 9	Total <i>n</i> = 177
<20	11 (50%)	9 (13%)	12 (18%)	2 (22%)	1 (11%)	35 (20%)
20–100	8 (35%)	28 (39%)	30 (46%)	2 (22%)	3 (33%)	71 (40%)
100–200	2 (9%)	9 (13%)	5 (8%)	2 (22%)	0 (0%)	18 (10%)
200–1000	0 (0%)	14 (19%)	15 (23%)	1 (11%)	1 (11%)	31 (17%)
>1000	1 (5%)	7 (10%)	1 (4%)	1 (11%)	3 (33%)	13 (7%)
Unknown	0 (0%)	5 (7%)	1 (3%)	1 (11%)	1 (11%)	9 (5%)

<sup>a</sup>In all tables, *n* = number of papers in the group.

(citations for these papers are included in the Supplemental Material for reference, and several that are particularly relevant are cited and summarized in the *Discussion*).

Among the remaining 268 papers from 2014 to 2017, 177 were coded for *study* type as empirical studies or program evaluation, while the remaining papers were program description only, theoretical, review, or other. To meet the call from Linn et al. (2015) for more empirical studies, and because the distinction between empirical and program evaluation was not always clear, we grouped the two categories that included analysis of data (empirical and program evaluation), terming the study type for these 177 papers as “data driven.”

In the remainder of the *Results* section, we present our analysis of these 177 papers (data-driven papers published from 2014 to 2017), because they had the most potential to provide descriptions as well as evidence-based claims that may be translatable to other programs in the field (see the Supplemental Material for a reference list of these 177 papers).

Of the 177 data-driven papers published during the 2014–2017 time period, more studies involving CUREs (*n* = 72; 41%) and UREs (*n* = 65; 37%) were published, compared with TRE studies (*n* = 22; 12%). The classification of “other” (*n* = 9; 5%) includes studies that did not fall into the CURE–URE–TRE coding scheme, whereas the “combination” classification (*n* = 9; 5%) includes publications that studied a mixture of two or more types of research experiences. Interestingly, of the 268 total papers from 2014 to 2017, TREs had a significantly greater percentage of data-driven papers (85%) than CUREs (64%) or UREs (77%; *p* = 0.027).

The majority of undergraduate and teacher research programs described had a disciplinary focus in biology (58%). Other fields include physical science (31%), earth and space science (11%), engineering (20%), mathematics (8%), computer science (7%), interdisciplinary STEM fields (17%), other (6%), and not stated (6%). The predominance of biological research was also the case for each of the three program types: CUREs (64%), UREs (57%), and TREs (45%). Note that multiple codes were permitted for disciplinary focus.

### Study Participants

The number of study participants reported in each paper was determined overall and compared across program types (Table 1). The overall data were clustered across categories of less than 20 (*n* = 35), 20–100 (*n* = 71), 100–200 (*n* = 18), 200–1000 (*n* = 31), more than 1000 (*n* = 13), and unknown (*n* = 9). The number of participants in each cluster was then broken down across different program types (Table 1). The majority of TRE programs involved fewer than 20 participants and

the majority of CURE and URE programs involved between 20 and 100 participants. There were 14 CURE and 15 URE studies that involved between 200 and 1000 participants. Of the studies that collected data from more than 1000 participants, most were CUREs (*n* = 7). The only TRE study of more than 1000 participants was also one of the largest studies among this data set (*n* = 10,468; Ragusa and Juarez, 2017). Papers reporting on TRE experiences primarily studied in-service teachers (*n* = 20); few studied preservice teachers (*n* = 0) or both (*n* = 2).

Note that these data were obtained by coding for total participant number reported. In most cases, the total number reported referred to the participant number in the particular research experience program and was the same as the number of study participants, but in some cases the total participant number referred to the total study participants in several different programs combined or programs from consecutive years or to a subset of program participants.

### Identification of Participants from Underrepresented Groups

A number of prior studies pointed to the importance of considering how participants from underrepresented groups experience CURE, URE, and TRE settings, as experiences and outcomes for underrepresented undergraduate students and teachers may differ from those of other program participants (Jones et al., 2010; Junge et al., 2010; Ovink and Veazey, 2011; Schwartz, 2012; Slovacek et al., 2012; Stevens and Hoskins, 2014; Linn et al., 2015). Thus, our analysis specifically sought mention of, and disaggregated outcomes for, participants from underrepresented groups, which was coded by author identification of study and program populations (underrepresented groups include women, persons of color, and persons with disabilities in science and engineering). Of the 177 studies, 114 (64%) failed to mention the involvement of teachers or undergraduate students from underrepresented groups (Table 2). While 47 studies (26%) identified the number of participants from underrepresented groups in the program and study, they did not report findings specific to these participants. Only 17 (10%) reported an intentional focus on participants from underrepresented groups. This general trend repeated across all categories of research experiences. Specifically, the majority of TRE studies did not mention underrepresented groups, and there were no TRE studies that focused explicitly on participants from underrepresented groups.

### Duration and Intensity of Research Experience

Because others have identified variable outcomes related to the duration and level of intensity of the research experience

TABLE 2. Number of studies by participant outcome data and program: underrepresented (UR) populations

Outcome data	TRE <i>n</i> = 22	CURE <i>n</i> = 72	URE <i>n</i> = 65	Combination <i>n</i> = 9	Other <i>n</i> = 9	Total <i>n</i> = 177
UR focus	0 (0%)	1 (1%)	15 (23%)	0 (0%)	1 (11%)	17 (10%)
UR identified	4 (18%)	13 (18%)	23 (35%)	3 (33%)	3 (33%)	47 (26%)
UR not mentioned	18 (82%)	58 (81%)	27 (42%)	6 (67%)	5 (56%)	114 (64%)

(Sadler *et al.*, 2010; Estrada *et al.*, 2011; Shaffer *et al.*, 2014; Corwin *et al.*, 2015; Linn *et al.*, 2015), we sought information on research duration and intensity. In our analysis scheme, we define low intensity as less than 10 hours of participation per week and high intensity as more than 10 hours of participation per week. Previously, Linn *et al.* (2015) reported that most CUREs involved lower-division students (first- and second-year undergraduates) and lasted one semester or less. Most UREs involve a full range of students and variable lengths of participation (Corwin *et al.*, 2015). In our analysis, we found similar trends over the past few years across CURE, URE, and TRE papers. Overall, only 7% of the 177 data-driven studies described research experiences longer than 1 year ( $n = 12$ ), and most CUREs and UREs lasted for one term (3 to 4 months;  $n = 31$ , 43%, and  $n = 24$ , 37%, respectively). While 15% ( $n = 10$ ) of UREs and 29% ( $n = 21$ ) of CUREs lasted for less than one term, most TREs in our study fell into this category ( $n = 14$ , 64%), which may result from teachers' participation during the summer (ranging from 1 week to 3 months). The percentage of TREs in this category was a statistically higher percentage than for UREs or CUREs ( $p = 0.0001$ ). Across all studies, a noteworthy proportion did not state the duration ( $n = 32$ , 18%) or intensity of the research experience ( $n = 58$ , 33%). Most CURE studies reported low-intensity research experiences ( $n = 50$ , 69%), but most URE and TRE studies reported intensity either as high ( $n = 25$ , 38%, and  $n = 11$ , 50%, respectively) or failed to state the intensity of the research experience ( $n = 29$ , 45%, and  $n = 8$ , 36% respectively;  $p = 0.0001$ ).

### Study Methodology

Prior studies highlighted the need to examine the impacts of research experience programs through a variety of data sources. For example, Sadler *et al.* (2010) found that most programs for

undergraduates and programs for teachers relied solely on participant self-reports of learning outcomes. Similarly, Linn *et al.*'s (2015) review found that more than half of the studies they examined relied solely on self-report surveys or interviews. From our analysis, we suggest that this call has begun to be answered, with 57% ( $n = 101$ ) of the 177 data-driven reports containing measures other than, or in addition to, self-report, a significant increase across CUREs, UREs, and TREs ( $p = 0.047$ ). The most substantial proportion were TRE studies, which more frequently used measures beyond self-report data ( $n = 16$ , 73%), compared with CURE ( $n = 42$ , 58%) or URE ( $n = 30$ , 46%) studies ( $p = 0.047$ ).

Sadler *et al.* (2010) state, "The most pressing issue that we have raised is the need for greater methodological diversity" (p. 253). Accordingly, we coded the sample for methodology and data sources, both quantitative and qualitative, as well as for research and evaluation purposes. Results (Table 3) show that many studies used a variety of data sources and approaches. While many studies continue to use quantitative self-report surveys (such as Likert-scale items), they use additional measures simultaneously. The use of qualitative self-report surveys (such as open-ended questions) to measure participant perspective is high among TRE programs ( $n = 14$ , 64%) compared with URE and CURE programs ( $n = 40$ , 56%, and  $n = 28$ , 43%, respectively), but not to a statistically significant degree ( $p = 0.164$ ). Among measures other than surveys, 23% ( $n = 41$ ) used institutional data, and 24% used a content or practice assessment ( $n = 42$ ). Methods other than surveys used across CURE, URE, and TRE studies varied; CURE studies most frequently used content/practice assessments ( $n = 27$ , 38%), URE studies most frequently used institutional data ( $n = 21$ , 32%), and TRE studies most frequently used interviews/focus groups ( $n = 12$ , 55%). Also note that many studies used a variety of methods, as

TABLE 3. Number of studies by data type and program

Data type <sup>a</sup>	TRE <i>n</i> = 22	CURE <i>n</i> = 72	URE <i>n</i> = 65	Combination <i>n</i> = 9	Other <i>n</i> = 9	Total <i>n</i> = 177
Self-report only	6 (27%)	30 (42%)	35 (54%)	4 (44%)	1 (11%)	76 (43%)
More than self-report	16 (73%)	42 (58%)	30 (46%)	5 (56%)	8 (89%)	101 (57%)
Institutional/extrinsic	0 (0%)	14 (19%)	21 (32%)	2 (22%)	3 (33%)	41 (23%)
Quant. participant <sup>b</sup>	12 (55%)	42 (58%)	37 (57%)	5 (56%)	5 (56%)	101 (57%)
Qual. participant <sup>b</sup>	14 (64%)	40 (56%)	27 (42%)	6 (67%)	1 (11%)	88 (50%)
Quant. faculty/mentor <sup>b</sup>	0 (0%)	4 (6%)	3 (5%)	1 (11%)	0 (0%)	8 (5%)
Qual. faculty/mentor <sup>b</sup>	1 (5%)	2 (3%)	2 (3%)	1 (11%)	0 (0%)	6 (3%)
Interview <sup>c</sup>	12 (55%)	8 (11%)	14 (22%)	0 (0%)	1 (11%)	35 (20%)
Content/practice	8 (36%)	27 (38%)	5 (8%)	1 (11%)	1 (11%)	42 (24%)
Other	8 (36%)	10 (14%)	6 (9%)	1 (11%)	1 (11%)	26 (15%)

<sup>a</sup>The 177 data-driven papers were independently coded for self-report only vs. more than self-report, and type of data collected (survey, interviews, content/practice assessment, other). A paper could receive multiple codes for type of data collected.

<sup>b</sup>Self-report survey, Quant, quantitative data (e.g., multiple choice, checkboxes) collected via survey; Qual, qualitative data (e.g., open-ended responses, essay questions) collected via survey.

<sup>c</sup>Interview includes focus groups.

**TABLE 4. Number of studies by design type and program**

Design type <sup>a</sup>	TRE n = 22	CURE n = 72	URE n = 65	Combination n = 9	Other n = 9	Total n = 177
One data point	0 (0%)	14 (19%)	20 (31%)	1 (11%)	1 (11%)	36 (20%)
Pre and post	8 (36%)	24 (33%)	15 (23%)	4 (44%)	3 (33%)	54 (31%)
Trajectory	8 (36%)	6 (8%)	6 (9%)	0 (0%)	2 (22%)	22 (12%)
Quasi-experimental	2 (9%)	12 (17%)	4 (6%)	0 (0%)	1 (11%)	19 (11%)
Longitudinal	0 (0%)	1 (1%)	4 (6%)	1 (11%)	1 (11%)	7 (4%)
Other	1 (5%)	1 (1%)	4 (6%)	0 (0%)	1 (11%)	7 (4%)

<sup>a</sup>One data point included pre or post only; trajectory included retrospective and/or multiple data points over time; quasi-experimental included comparison of 2+ sample conditions; longitudinal included tracking individuals over time.

the proportions add up to far beyond 100% for each program type and across specific data sources.

A final coding scheme for methodology included categorizing different study designs with respect to when data were collected (Table 4). While most studies used a pre/post measure (n = 54, 31%), a number were of more complex design, such as multiple data points throughout the program (n = 22, 12%), quasi-experimental (n = 19, 11%), and longitudinal (n = 7, 4%). A proportion of the TRE studies (n = 8, 36%) used either multiple data points over time or a trajectory/retrospective study design, for example, as a method of classroom follow-up after the research experience.

**Measured Outcomes**

The three previous reviews presented well-organized lists of reported measured outcomes. Authors included similar reported outcomes in the coding scheme used here (Table 5). Across all programs, the most frequently reported outcomes measured were science practices (n = 67, 38%), laboratory skills (n = 61, 35%), disciplinary content knowledge (n = 60, 34%), and confidence (n = 55, 31%). These reports indicate

positive gains with respect to the measured outcomes. Among the TRE studies, these four outcomes were relatively sparse, with “impacts on classroom practice” being the most targeted outcome (n = 13, 60%). Surprisingly, only 23% (n = 5) of TRE studies focus on K–12 student outcomes, which is a stated long-term goal for most TRE programs (see *Discussion*).

**Author-Identified Theoretical Framework**

Linn et al. (2015) proposed the application of the knowledge integration framework for research on undergraduate research experiences, which provided a guide for their study and enabled them to identify gaps and challenges in the field. They suggested that reports of research experiences need to more frequently identify the theoretical foundation that framed their studies as well as programs. Of note in the data-driven studies in the current review, only 37% (n = 66) explicitly identified a framework (theoretical, conceptual, or learning theory) that informed their approach (Table 6). Author-identified theoretical frameworks (i.e., the structure that supports the theoretical basis of a research study; see also Sutton and Staw, 1995) included cognitive apprenticeship;

**TABLE 5. Number of studies by reported measured outcomes and program (2014–2017)**

Measured outcomes <sup>a</sup>	TRE n = 22	CURE n = 72	URE n = 65	Combination n = 9	Other n = 9	Total n = 177
Not stated	0 (0%)	1 (1%)	3 (5%)	0 (0%)	1 (11%)	5 (3%)
Performance <sup>b</sup>	0 (0%)	13 (18%)	13 (20%)	1 (11%)	1 (11%)	28 (16%)
Content knowledge <sup>c</sup>	5 (23%)	35 (49%)	17 (26%)	3 (33%)	0 (0%)	60 (34%)
NOS	5 (23%)	12 (17%)	10 (15%)	0 (0%)	0 (0%)	27 (15%)
Persistence <sup>b</sup>	0 (0%)	12 (17%)	24 (37%)	3 (33%)	3 (33%)	42 (24%)
Science practices	5 (23%)	33 (46%)	25 (38%) <sup>d</sup>	2 (22%)	2 (22%)	67 (38%) <sup>d</sup>
Lab skills	1 (5%)	36 (50%) <sup>d</sup>	20 (31%)	4 (44%) <sup>d</sup>	0 (0%)	61 (35%)
21st-century skills	1 (5%)	15 (21%)	10 (15%)	1 (11%)	0 (0%)	27 (15%)
Self-efficacy	9 (41%)	13 (18%)	15 (23%)	2 (22%)	4 (44%) <sup>d</sup>	43 (25%)
Confidence	4 (18%)	26 (36%)	19 (29%)	3 (33%)	3 (33%)	55 (31%)
Attitudes/interest <sup>b</sup>	4 (18%)	16 (22%)	19 (29%)	4 (44%) <sup>d</sup>	1 (11%)	44 (25%)
Teacher identity	1 (5%)	2 (3%)	0 (0%)	1 (11%)	0 (0%)	4 (2%)
Scientist identity	2 (9%)	8 (11%)	9 (14%)	1 (11%)	1 (11%)	21 (12%)
Classroom practice	13 (60%) <sup>d</sup>	1 (1%)	0 (0%)	1 (11%)	2 (22%)	17 (10%)
K–12 outcomes	5 (23%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	5 (3%)
Perceptions	4 (18%)	2 (3%)	0 (0%)	0 (0%)	1 (11%)	7 (4%)
Awareness <sup>b</sup>	3 (14%)	0 (0%)	12 (18%)	1 (11%)	0 (0%)	16 (9%)
Leadership	0 (0%)	0 (0%)	3 (5%)	1 (11%)	0 (0%)	4 (2%)

<sup>a</sup>Performance includes course grades and/or grade point average. Perceptions refers to teachers and teaching. NOS, nature of science.

<sup>b</sup>Pertains to STEM careers.

<sup>c</sup>Pertains to science content discipline knowledge.

<sup>d</sup>Indicates the most frequent category in the group.



**TABLE 6. Papers with at least one author-identified theoretical framework**

Framework	Number (%)
TRE ( $n = 22$ )	13 (59)
CURE ( $n = 72$ )	23 (32)
URE ( $n = 65$ )	25 (38)
Combination ( $n = 9$ )	3 (33)
Other ( $n = 9$ )	2 (22)
Total ( $n = 177$ )	66 (37)

situated learning theory; communities of practice, identity, and agency; and social cognitive career theory. Interestingly, a much higher proportion of TRE studies ( $n = 13$ , 59%) describe a framework compared with CURE ( $n = 23$ , 32%) and URE ( $n = 25$ , 38%) studies, but not to a statistically significant degree ( $p = 0.072$ ).

### DISCUSSION, IMPLICATIONS, AND DIRECTIONS FOR FUTURE RESEARCH

This study provides updates to calls sounded by Sadler *et al.* (2010), Corwin *et al.* (2015), Linn *et al.* (2015), and NASEM (2017). We focus on our key research questions sequentially in the following sections: 1) How has the research literature about science research experiences for teachers and undergraduates changed over the past decade? 2) What is the involvement of, and impact on, participants from underrepresented groups in science research experiences? 3) How are science research experiences impacting teachers and science teaching? 4) How are studies of research experience programs being conducted and what type of data sources are in use? Finally, we address the need for synergistic approaches between science research experts and science education experts when designing and conducting science research experiences. Our findings reveal trends as well as limitations in the recent research literature in several areas. We address critical points for these findings by discussing exemplary studies and outline recommendations for future action.

#### How Has the Research Literature about Science Research Experiences Changed over the Past Decade?

Sadler *et al.* (2010) examined 53 papers written between 1961 and 2008, 22 of which were publications about apprenticeship research experiences for undergraduates, 20 about secondary education students, and 11 about research programs for teachers. In comparison, our analysis of 177 data-driven papers written between 2014 and 2017, found that 72 studies focused on CUREs, 65 focused on UREs, 22 studies focused on TREs, and 18 studies focused on a combination of these groups or were grouped as “other.” Thus, there is a substantial increase overall in the number of studies involving science research experiences, especially for undergraduate research. While the overall growth is contributing to our understanding of the nature and impact of such research experiences, we note that the decreased proportion of studies (i.e., from 21 to 12%) focused specifically on science teacher development raises questions about the direction of the field. Involving as many learners as possible in research experiences is encouraging. However, as teacher educators, we need to continue improving

how we prepare teachers to engage learners in authentic science practices. Given the increase in CUREs and UREs, there is an opportunity for enhancing teacher preparation programs. Purposefully including courses with embedded research experiences and other undergraduate research opportunities within teacher preparation could provide essential contextual and disciplinary knowledge related to scientific practices and the scientific community. *Facilitating the transfer of science research experiences to inquiry-based science pedagogy could be targeted as part of CURE and URE programs, as well as during pedagogy and practicum courses for future teachers who participate in CURE or URE programs. In these ways, undergraduate research experiences could more purposefully address the needs of future teachers.*

#### What Is the Involvement of, and Impact on, Participants from Underrepresented Groups in Science Research Experiences?

Science research experiences have been shown in earlier studies to positively impact the achievement and STEM interests of participants from underrepresented groups (Nagda *et al.*, 1998; Davis, 1999; Sadler *et al.*, 2010; NASEM, 2017). Nagda and colleagues (1998) reported positive impacts on retention of undergraduates in science majors, particularly among African American students. Linn *et al.* (2015) concluded that “studies that analyze benefits for subgroups of students could also help those designing research experiences address the unique interests and aspirations of individuals and groups” and that “findings about successful experiences need replication and extension, particularly for students from non-dominant cultures” (2015, p. 5). Despite these findings and the intentions of funding agencies to support programs that address the needs of people of color and other underserved populations in STEM fields, we found that the majority of studies about CUREs, UREs, and TREs published between 2014 and 2017 fail to report demographic data that would identify the proportion of participants from underrepresented populations (64%), or when this information is reported, fail to disaggregate outcomes specifically for underrepresented students and/or teachers (26%). While many of these reports may mention increasing underrepresented populations in STEM as rationale for their program, most do not purposely target these groups or report relevant outcomes. Of the 17 papers (10%) that clearly targeted underrepresented groups and included disaggregated data for these populations, 15 were URE studies. The underrepresentation of such papers in the CURE literature reviewed here may reflect the perspective that CURE curricula are intended to be designed for all students. However, in a 2018 CURE study about the well-known Freshman Research Initiative, Corwin *et al.* (2018) call for more research on CUREs that report on outcomes for underrepresented groups, in order to understand how these research experiences may differ in their impact across groups.

Nine papers from our review of recent literature stood out as not only involving data analysis for underrepresented groups, but presented findings that we consider particularly relevant (Miranda and Damico, 2015; Robnett *et al.*, 2015; Shapiro *et al.*, 2015; Griffeth *et al.*, 2016; Haeger and Fresquez, 2016; Remich *et al.*, 2016; Carpi *et al.*, 2017; Katz *et al.*, 2017; Ghee *et al.*, 2018). A particularly noteworthy example is from Carpi *et al.* (2017), who suggest the value of an extended duration of

mentorship during URE participation (1–3 years), especially for participants from underrepresented groups, for increasing graduation rates and the number of students earning advanced degrees. In addition, Haeger and Fresquez (2016) explore research duration and the impact of socioemotional and culturally relevant mentoring, reporting increased grade point averages and level of independence in research. Ghee *et al.* (2018) focuses on professional development activities for students at minority-serving institutions, closely tracking the success of these students through career pathway measures. Finally, a comparison study by Katz *et al.* (2017) convincingly demonstrates STEM persistence for female students of color in particular.

Very recent work beyond the inclusion criteria for our current study offers additional insights into impacts of science research experiences on underrepresented minority (URM) participants. A longitudinal study by Hernandez *et al.* (2018) that focused primarily on URMs (94%), women (74%), and biology majors (72%) found that the duration and intensity level of a research experience impacted URM persistence in STEM, stating that undergraduate research was impactful only if students engaged in research for at least 10 hours/week for two or more semesters. Of the 72 CURE papers in our review of the literature, only 11 papers (15%) report a duration of more than one semester and only six papers (8%) report high-intensity level (10+ hours/week); no papers fall into both categories. Additional studies focused on the impacts of undergraduate research on URM persistence in STEM suggest that CUREs alone may not be able to offer the amount and depth of mentoring needed to support underrepresented students to stay in STEM fields, but may be useful as a stepping stone to a high-impact URE (Fuchs *et al.*, 2016; Tootle *et al.*, 2019).

We join others in reiterating the call for more studies that 1) purposefully target a diverse participant sample and 2) rigorously collect, analyze, and report data that reflect outcomes that may be unique for women and participants of color. Such studies will lead to increasingly better understanding of new approaches to undergraduate and teacher course and programs that result in enhanced representation in STEM fields. *If we are to make strides toward equity and diversity in STEM fields, researchers and program developers alike need to intentionally develop programs that not only attract and facilitate diverse participation, but also consider specific attention toward data collection, analysis, and reporting that reflects this goal.*

### How Are Science Research Experiences Impacting Teachers and Science Teaching?

A primary goal of research experiences for teachers is “to equip teachers with an understanding of and a capability to conduct scientific research that will transfer to their science classrooms” (Sadler *et al.*, 2010, p. 242). The literature also shows the importance of program elements that provide specific and supported opportunities for teachers to translate their research experiences into classroom instruction (Schwartz and Crawford, 2004; Sadler *et al.*, 2010). The current review demonstrates that outcomes that would indicate a transfer of science research knowledge to the science classroom (especially K–12 student learning outcomes) are *not* being measured sufficiently. Also, there has been insufficient attention to the outcomes that may provide insight into the success of these experiences, such as teacher identity or perceptions of the teaching profession.

While it is understandable that CURE and URE programs prioritize other goals (e.g., lab skills, science practices, attitudes), we emphasize to researchers and program developers the importance of acknowledging that teachers (preservice or in-service) are researchers with similar immediate outcomes to those students who participate in CUREs or UREs, but different professional interests, career aspirations, and long-term outcomes. *To measure results pertaining specifically to the K–12 classroom, research questions focused on assessment in this area must be developed and recognized as valid ways to measure the success of a thoughtfully planned and implemented research experience for teachers.*

Many studies that focus on TREs are of particular note, including papers published between 2007 and 2013 that were not included in previous reviews. We list several here with their alignment toward specific aspects of the CARET model (Figure 1). In particular, we specifically highlight TRE studies associated with communities of practice (Feldman *et al.*, 2009; Alkather and Dolan, 2014; McLaughlin and MacFadden, 2014; Salto *et al.*, 2014; Peters-Burton *et al.*, 2015; Luera and Murray, 2016; Southerland *et al.*, 2016); PLCs (Enderle *et al.*, 2014; Miranda and Damico, 2015; Southerland *et al.*, 2016); translation to practice (Pop *et al.*, 2010; Herrington *et al.*, 2011; McLaughlin and MacFadden, 2014; Amolins *et al.*, 2015; Godin *et al.*, 2015; Miranda and Damico, 2015; Blanchard and Sampson, 2017; Hardré *et al.*, 2017; Ragusa and Juarez, 2017); and K–12 assessments (Silverstein *et al.*, 2009; Yelamarthi *et al.*, 2013).

Particularly noteworthy for both undergraduate and teacher research is a study by Hanauer *et al.* (2017) that found measures of project ownership, scientific community values, science identity, and science networking reflected persistence in science. Some of these components can be addressed by using an approach such as that of Feldman *et al.* (2009), wherein a research group model is applied to the research apprenticeship, assigning specific roles to students, such as novice researcher, proficient technician, or knowledge producer. These align with a community of practice component. Indeed, reasons teachers give for participating in a research experience or selecting particular programs, can influence their outcomes (Enderle *et al.*, 2014). Enderle and colleagues (2014) found that those teachers looking to reform their instruction chose and showed considerable impacts from research experience programs explicitly designed for pedagogical development. Likewise, Southerland *et al.* (2016) found that providing personal relevance and social engagement in the research context increased investment and that participating in a research experience shapes science teaching beliefs, which in turn influences practice. Similarly, Miranda and Damico (2015) found that a summer TRE followed by an academic year-long PLC can help teachers to shift their beliefs surrounding pedagogical approaches; however, documentation of classroom practices that evidence this shift is limited. Such studies provide in-depth insight into the learning outcomes and impacts of research experiences that can serve as a model for future studies, particularly for teacher researchers.

### How Are Studies of Research Experience Programs Being Conducted and What Type of Data Sources Are in Use?

While the approach of using self-reported outcomes has been valuable for understanding undergraduate or teacher researcher perspectives for many years, as well as being an efficient means

of engaging in evaluation of particular programs, our findings suggest a current trend toward additional measures that hold promise for the development of a more in-depth understanding of CURE, URE, and TRE outcomes. First, researchers have developed many innovative instruments and approaches in recent years from which other studies could benefit. These instruments also inspire the development of new ways for measuring outcomes of research experiences. Papers of note in this area include Feldman *et al.*, 2009; Grove *et al.*, 2009; Kerlin, 2013; Yelamarthi *et al.*, 2013; Miranda and Damico, 2015; Zhou *et al.*, 2015; Felzien, 2016; Hanauer *et al.*, 2017; and Harsh *et al.*, 2017.

Second, a variety of novel methodologies or study designs were used in many recent papers, including those that did comparative or longitudinal studies, used multiple instruments, or performed analysis to determine a causal relationship. These papers include Silverstein *et al.*, 2009; Bahbah *et al.*, 2013; Saka, 2013; Enderle *et al.*, 2014; Yaffe *et al.*, 2014; Peters-Burton *et al.*, 2015; Robnett *et al.*, 2015; Shapiro *et al.*, 2015; Thompson *et al.*, 2015; Luera and Murray, 2016; Staub *et al.*, 2016; Stephens *et al.*, 2016; and Mader *et al.*, 2017. In particular, *future studies with designs that can shed light on correlational and causal relationships between factors involved in TREs, the PLC supporting the translation to practice, and longer-term teacher and student outcomes have great potential to advance the field* (Figure 1, CARET model).

A study conducted by Grove *et al.* (2009) serves as an example of how a variety of qualitative tools, including pre/post interviews, can be effectively used to test outcomes of general target goals such as inquiry, nature of science, experimental design, process skills, and communicating about science. Kerlin's (2013) study used observations, email solicitation of teacher ideas, and field notes in an innovative way, identifying critical factors toward incorporation of inquiry learning, including collaboration of STEM and education faculty with teacher participants and sustained support in the classroom. Noteworthy again, Hanauer *et al.*'s (2017) comprehensive report on the SEA-PHAGES program demonstrates an exciting array of outcomes with their use of the Psychometric Persistence in the Sciences survey, including project ownership (with content and emotion categories), self-efficacy, science identity, scientific community values, and networking, each measuring a psychological component that correlates strongly with a student's intention to continue in science. Finally, Klein-Gardner *et al.* (2012) studied a program in which teachers used a challenge-based curriculum (Legacy Cycle), that incorporated real-world contexts and interdisciplinary approaches to expose K–12 students to engineering, leading to a high level of engagement and participation among students in the classroom.

One of the most extensive studies among the collection was a 7-year longitudinal study involving 10,468 participants (Ragusa and Juarez, 2017). This work examined the impact of two TRE programs, reporting on the combined results of both programs on in-service teachers and their students, using an observational rubric to examine teaching performance and a Science Teaching Efficacy Beliefs Instrument (Enochs and Riggs, 1990) and K–12 student science reading and concept inventory assessments. Findings support the hypothesis that a teacher intervention results in student outcome gains. While

this longitudinal study has a uniquely large sample, such research is essential for connecting research experience models, pedagogical impacts, and student impacts.

Enderle *et al.* (2014) is one of the few studies that used a mixed-methods approach to compare impacts related to two distinct summer research models for teachers: one pedagogically focused and the other science research focused. Over a 5-year period, they measured science teachers' self-efficacy, pedagogical discontentment, teachers' beliefs about teaching and learning, and contextual beliefs about teaching science. Results indicated both types of research experience programs positively impacted participants' beliefs, with the teachers from the pedagogically focused program showing greater gains compared with the science research-focused group. In addition, teachers in the pedagogically focused program were successful in reforming their instructional practice; whereas the science research-focused teachers were not. This study shows the importance of multiple data sources, collected over multiple years of program implementation, for identifying consistent and overarching trends concerning program impacts on participants' beliefs and potential for translation to classroom practice.

The recent NASEM report (2017) provides detailed recommendations for methodology that would be useful for surfacing causal relationships among components of UREs (chap. 7, pp. 163–180); most of these recommendations also apply to research on TREs.

### **Suggestions for Increasing Synergistic Approaches between Science Research Experts and Science Education Experts When Designing and Conducting Science Research Experiences**

Among the publications examined in this literature review, the occurrence of productive research efforts through collaborations between STEM-disciplinary faculty/industry researchers and education faculty is rare; however, we identified several relevant efforts. Boesdorfer and Asprey (2017) demonstrated that research experiences can be science education research experiences and found a distinct difference in the documented outcomes between science research experiences and science education research experiences. A recent National Academies report (NASEM, 2017) concluded that URE designers were not taking full advantage of the education literature in the design, implementation, and evaluation of UREs. The report recommended that those with science education expertise should conduct studies in collaboration with URE program directors to address how program components impact outcomes. *We echo this call and encourage future researchers to form authentic collaborative research partnerships among content area experts and pedagogical experts. We also note that such collaborations could be significant for further understanding effective communities of practice and PLCs outlined in the CARET model.*

### **CONCLUSIONS**

This review advances our collective understanding of the current state regarding science research experiences for teachers and undergraduates. Researchers have made advances in studies of CURE and URE programs. These types of programs include future teachers, and we hope to raise awareness within these programs that these undergraduates may benefit further from pedagogical reflection, as many choose a career in



teaching in K–12 schools or in higher education. Reports of TRE programs, while decreasing in frequency, are providing some glimpse into instructional impacts, especially within the past 4 years. We call for more studies with such focus in order to correlate program features with instructional outcomes.

Moreover, we found that, despite funding agencies' explicit goals for increasing representation in STEM fields, few recent studies describe how they select participants or provide additional insights into outcomes of research experiences for people of color and underserved groups. Intentional selection, design, and support are needed for STEM research experience programs in general, and science teachers in particular. Finally, we have highlighted exemplary studies that used methodological designs and multiple data sources that lend confidence to the research findings and effectiveness of the programs. Again, we recommend continued advances in research approaches that provide insights into teacher development, instructional reform, and student impacts.

We offer the CARET model as a conceptual framework that can inform program development and future studies on teacher STEM research experiences. The model provided an initial perspective that guided our coding scheme. More work is needed to test components of this preliminary model and potentially expand it to incorporate finer-grained features. More specific participation and collection of data from underrepresented populations and development of practical, valid, and reliable instruments for measuring K–12 teaching practices and student learning outcomes will contribute to a more accurate view of the success of the TRE programs. Finally, we encourage authentic collaboration between content experts and pedagogical experts to forward these goals.

## ACKNOWLEDGMENTS

This work was made possible through funding for the CARET group to attend several meetings and conferences provided through a 100Kin10 collaboration grant, funding from the APLU Network of STEM Education Centers (NSEC), in-kind support from the MJ Murdock Charitable Trust, and several awards from the NSF including the following programs: Robert Noyce Teacher Scholarship, Research Experiences for Teachers, Graduate Research Fellowship Program, and Improving Undergraduate Science Education. This material is based upon work supported by the NSF under Grant Nos. (1524832, 1340042, 1540826, 1542471, 1106400, and 1712001). We thank these organizations for their generous support. We also extend a special thank you for the guidance and contribution provided by Bruce Johnson of the University of Arizona.

## REFERENCES

- Alkaker, I., & Dolan, E. L. (2014). Integrating research into undergraduate courses: Current practices and future directions. In Sunal, D., Sunal, C., Zollman, D., Mason, C., & Wright, E. (Eds.), *Research in science education: Research based undergraduate science teaching* (pp. 403–430). Charlotte, NC: Information Age.
- Amolins, M. W., Ezrailson, C. M., Pearce, D. A., Elliott, A. J., & Vitiello, P. F. (2015). Evaluating the effectiveness of a laboratory-based professional development program for science educators. *Advances in Physiology Education*, 39(4), 341–351. <https://doi.org/10.1152/advan.00088.2015>
- Astin, A. W., & Oseguera, L. (2004). The declining "equity" of American higher education. *Review of Higher Education*, 27(3), 321–341. <https://doi.org/10.1353/rhe.2004.0001>
- Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., ... & Dolan, E. L. (2014). Assessment of course-based undergraduate research experiences: A meeting report. *CBE—Life Sciences Education*, 13(1), 29–40. <https://doi.org/10.1187/cbe.14-01-0004>
- Bahbah, S., Golden, B. W., Roseler, K., Elderle, P., Saka, Y., & Southerland, S. A. (2013). The influence of RET's on elementary and secondary grade teachers' views of scientific inquiry. *International Education Studies*, 6(1), 117–131. <https://doi.org/10.5539/ies.v6n1p117>
- Blanchard, M. R., & Sampson, V. D. (2017). Fostering impactful research experiences for teachers (RETs). *Eurasia Journal of Mathematics, Science and Technology Education*, 14(1), 447–465. <https://doi.org/10.12973/ejmste/80352>
- Boesdorfer, S. B., & Asprey, L. M. (2017). Exploratory study of the teaching practices of novice science teachers who participated in undergraduate science education research. *Electronic Journal of Science Education*, 2(3), 21–45. Retrieved August 5, 2019, from <https://files.eric.ed.gov/fulltext/EJ1187994.pdf>
- Campbell, A. M. (2002). *The influences and factors of an undergraduate research program in preparing women for science careers* (Unpublished doctoral dissertation). Texas Tech University, Lubbock, TX.
- Capps, D., & Crawford, B. (2013). Inquiry-based instruction and teaching about nature of science: Are they happening? *Journal of Science Teacher Education*, 24(3), 497–526. <https://doi.org/10.1007/s10972-012-9314-z>
- Carpi, A., Ronan, D. M., Falconer, H. M., & Lents, N. H. (2017). Cultivating minority scientists: Undergraduate research increases self-efficacy and career ambitions for underrepresented students in STEM. *Journal of Research in Science Teaching*, 54(2), 169–194. <https://doi.org/10.1002/tea.21341>
- Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. *CBE—Life Sciences Education*, 14(1), es1. <https://doi.org/10.1187/cbe.14-10-0167>
- Corwin, L. A., Runyon, C. R., Ghanem, E., Sandy, M., Clark, G., Palmer, G. C., ... & Dolan, E. L. (2018). Effects of discovery, iteration, and collaboration in laboratory courses on undergraduates' research career intentions fully mediated by student ownership. *CBE—Life Sciences Education*, 17(2), ar20. <https://doi.org/10.1187/cbe.17-07-0141>
- Crawford, B. (2014). From inquiry to scientific practices in the science classroom. In Lederman, N., & Abell, S. (Eds.), *Handbook of research on science education* (Vol. 2, pp. 515–541). New York: Routledge.
- Davis, D. D. (1999, April). The research apprenticeship program: Promoting careers in biomedical sciences and the health professions for minority populations. Paper presented at: annual meeting of the American Educational Research Association (Montreal, Ontario). Retrieved August 5, 2019, from <https://files.eric.ed.gov/fulltext/ED430494.pdf>
- Dogan, S., Pringle, R., & Mesa, J. (2016). The impacts of professional learning communities on science teachers' knowledge, practice and student learning: A review. *Professional Development in Education*, 42(4), 569–588. <http://doi.org/10.1080/19415257.2015.1065899>
- Duschl, R., Schweingruber, H., & Shouse, A. (Eds.) (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press.
- Enderle, P., Dentzau, M., Roseler, K., Southerland, S., Granger, E., Hughes, R., ... & Saka, Y. (2014). Examining the influence of RETs on science teacher beliefs and practice. *Science Teacher Education*, 98(6), 1077–1108. <https://doi.org/10.1002/sce.21127>
- Enochs, L., & Riggs, I. (1990). Further development of an elementary science teaching efficacy belief instrument: A preservice elementary scale. *School Science and Mathematics*, 90(8), 694–706. <https://doi.org/10.1111/j.1949-8594.1990.tb12048.x>
- Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, P. W. (2011). Toward a model of social influence that explains minority student integration into the scientific community. *Journal of Educational Psychology*, 103(1), 206–222. <https://doi.org/10.1037/a0020743>
- Feldman, A., Divoll, K., & Rogan-Klyve, A. (2009). Research education of new scientists: Implications for science teacher education. *Journal of Research in Science Teaching*, 46(4), 442–459. <https://doi.org/10.1002/tea.20285>
- Felzien, L. K. (2016). Integration of a zebrafish research project into a molecular biology course to support critical thinking and course content



- goals. *Biochemistry and Molecular Biology Education*, 44(6), 565–573. <https://doi.org/10.1002/bmb.20983>
- Fuchs, J., Kouyate, A., Kroboth, L., & McFarland, W. (2016). Growing the pipeline of diverse HIV investigators: The impact of mentored research experiences to engage underrepresented minority students. *AIDS and Behavior*, 20(2), 249–257. <https://doi.org/10.1007/s10461-016-1392-z>
- Ghee, M., Keels, M., Collins, D., Neal-Spence, C., & Baker, E. (2018). Fine-tuning summer research programs to promote underrepresented students' persistence in the STEM pathway. *CBE—Life Sciences Education*, 15(3), ar28. <https://doi.org/10.1187/cbe.16-01-0046>
- Godin, E. A., Wormington, S. V., Perez, T., Barger, M. M., Snyder, K. E., Richman, L. S., ... & Linnenbrink-Garcia, L. (2015). A pharmacology-based enrichment program for undergraduates promotes interest in science. *CBE—Life Sciences Education*, 14(4), ar40. <https://doi.org/10.1187/cbe.15-02-0043>
- Griffeth, N., Batista, N., Grosso, T., Arianna, G., Bhatia, R., Boukerche, F., ... & Krynski, K. (2016). An undergraduate research experience studying ras and ras mutants. *IEEE Transactions on Education*, 59(2), 91–97. <https://doi.org/10.1109/TE.2015.2450683>
- Grove, C. M., Dixon, P. J., & Pop, M. M. (2009). Research experiences for teachers: Influences related to expectancy and value of changes to practice in the American classroom. *Professional Development in Education*, 35(2), 247–260. <https://doi.org/10.1080/13674580802532712>
- Haeger, H., & Fresquez, C. (2016). Mentoring for inclusion: The impact of mentoring on undergraduate researchers in the sciences. *CBE—Life Sciences Education*, 15(3), ar36. <https://doi.org/10.1187/cbe.16-01-0016>
- Hanauer, D. I., Graham, M. J., Betancur, L., Bobrownicki, A., Cresawn, S. G., Garlena, R. A., ... & Hatfull, G. F. (2017). An inclusive research education community (iREC): Impact of the SEA-PHAGES program on research outcomes and student learning. *Proceedings of the National Academy of Sciences USA*, 114(51), 13531–13536. <https://doi.org/10.1073/pnas.1718188115>
- Hardré, P. L., Ling, C., Shehab, R. L., Nanny, M. A., Nollert, M. U., Refai, H., ... & Huang, S. M. (2017). Situating teachers' developmental engineering experiences in an inquiry-based, laboratory learning environment. *Teacher Development*, 21(2), 243–268. <https://doi.org/10.1080/13664530.2016.1224776>
- Harsh, J., Esteb, J. J., & Maltese, A. V. (2017). Evaluating the development of chemistry undergraduate researchers' scientific thinking skills using performance data: First findings from the performance assessment of undergraduate research (PURE) instrument. *Chemistry Education Research and Practice*, 18(3), 472–285. <https://doi.org/10.1039/C6RP00222F>
- Hasson, E., & Yarden, A. (2012). Separating the research question from the laboratory techniques: Advancing high-school biology teachers' ability to ask research questions. *Journal of Research in Science Teaching*, 49(10), 1296–1320. <https://doi.org/10.1002/tea.21058>
- Hernandez, P. R., Woodcock, A., Estrada, M., & Schultz, P. W. (2018). Undergraduate research experiences broaden diversity in the scientific workforce. *BioScience*, 68(3), 204–211. <https://doi.org/10.1093/biosci/bix163>
- Herrington, D. G., Yezierski, E. J., Luxford, K. M., & Luxford, C. J. (2011). Target inquiry: Changing chemistry high school teachers' classroom practices and knowledge and beliefs about inquiry instruction. *Chemistry Education Research and Practice*, 12(1), 74–84. <https://doi.org/10.1039/C1RP90010B>
- Hoxby, C. M., & Avery, C. (2013). The missing "one-offs": The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity*, 2013(1), 1–65. Retrieved December 1, 2018, from Project MUSE database. <https://doi.org/10.1353/eca.2013.0000>
- Ignite. (2018). *Summer fellowships*. Retrieved December 1, 2018, from [www.igniteducation.org](http://www.igniteducation.org)
- Jones, M. T., Barlow, A. E. L., & Villarejo, M. (2010). Importance of undergraduate research for minority persistence and achievement in biology. *Journal of Higher Education*, 81(1), 82–115. <https://doi.org/10.1353/jhe.0.0082>
- Junge, B., Quiñones, C., Kakietek, J., Teodorescu, D., & Marsteller, P. (2010). Promoting undergraduate interest, preparedness, and professional pursuit in the sciences: An outcomes evaluation of the SURE program at Emory University. *CBE—Life Science Education*, 9(2), 119–132. <https://doi.org/10.1187/cbe.09-08-0057>
- Katz, L. A., Aloisio, K. M., Horton, N. J., Ly, M., Pruss, S., Queeney, K., ... & DiBartolo, P. M. (2017). A program aimed toward inclusive excellence for underrepresented undergraduate women in the sciences. *CBE—Life Sciences Education*, 16(1), ar11. <https://doi.org/10.1187/cbe.16-01-0029>
- Kerlin, S. (2013). Professional development strategies that promote science inquiry teaching and learning. *Kentucky Journal of Excellence in College Teaching and Learning*, 10(6), 74–87. Retrieved August 5, 2019, from <https://encompass.eku.edu/kjlect/vol10/iss2012/6>
- Klein-Gardner, S. S., Johnston, M. E., & Benson, L. (2012). Impact of RET teacher-developed curriculum units on classroom experiences for teachers and students. *Journal of Pre-College Engineering Education Research*, 2(2), 21–35. <https://doi.org/10.5703/1288284314868>
- Kowalski, M., Schwartz, R., Sydlik, M. A., Bryson, T., & Shi, R. (2018, October). Challenges associated with reaching intercoder agreement across a research team. Paper presented at: annual conference of the American Evaluation Association (Cleveland, OH).
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Linn, M. C., Palmer, E., Baranger, A., Gerard, E., & Stone, E. (2015). Undergraduate research experiences: Impacts and opportunities. *Science*, 347(6222), 627–633. <https://doi.org/10.1126/science.1261757>
- 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), 303–313. <https://doi.org/10.5408/14-064.1>
- Mader, C. M., Beck, C. W., Grillo, W. H., Hollowell, G. P., Hennington, B. S., Staub, N. L., ... & Bradford, C. (2017). Multi-institutional, multidisciplinary study of the impact of course-based research experiences. *Journal of Microbiology & Biology Education*, 18(2), 1–11. <https://doi.org/10.1128/jmbe.v18i2.1317>
- McLaughlin, C. A., & MacFadden, B. J. (2014). At the elbows of scientists: Shaping science teachers' conceptions and enactment of inquiry-based instruction. *Research in Science Education*, 44(6), 927–947. <https://doi.org/10.1007/s11165-014-9408-z>
- Miranda, R. J., & Damico, J. B. (2015). Changes in teachers' beliefs and classroom practices concerning inquiry-based instruction following a year-long RET-PLC program. *Science Educator*, 24(1), 23–35. Retrieved August 5, 2019, from <https://files.eric.ed.gov/fulltext/EJ1069987.pdf>
- Nagda, B. A., Gregerman, S. R., Jonides, J., von Hippel, W., & Lerner, J. S. (1998). Undergraduate student-faculty research partnerships affect student retention. *Review of Higher Education*, 22(1), 55–72.
- National Academies of Sciences, Engineering, and Medicine. (2017). *Undergraduate research experiences for STEM students: Successes, challenges, and opportunities*. Washington, DC: National Academies Press. <https://doi.org/10.17226/24622>
- National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. Washington, DC: National Academies Press. Retrieved August 5, 2019, from <https://doi.org/10.17226/12984>
- National Research Council (NRC). (1996). *National science education standards*. Washington, DC: National Academies Press. Retrieved August 5, 2019, from [www.nap.edu/download/4962](http://www.nap.edu/download/4962)
- NRC. (2012). *A framework for K–12 science education: Practices, cross-cutting concepts, and core ideas*. Washington, DC: National Academies Press. Retrieved August 5, 2019, from [www.nap.edu/download/13165](http://www.nap.edu/download/13165)
- NRC. (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press. Retrieved August 5, 2019, from [www.nap.edu/download/18290](http://www.nap.edu/download/18290)
- National Science Foundation. (2019). *Research experiences for undergraduates (REU)*. Retrieved August 5, 2019, from [www.nsf.gov/funding/pgm\\_summ.jsp?pins\\_id=5517&from=fund](http://www.nsf.gov/funding/pgm_summ.jsp?pins_id=5517&from=fund)
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press. Retrieved August 5, 2019, from [www.nextgenscience.org/](http://www.nextgenscience.org/)
- Ovink, S. M., & Veazey, B. D. (2011). More than "getting us through": A case study in cultural capital enrichment of underrepresented minority undergraduates. *Research in Higher Education*, 52(4), 370–394. <https://doi.org/10.1007/s11162-010-9198-8>
- Partnership for 21st Century Learning. (2018). *Framework for 21st century learning*. Retrieved from [www.p21.org/our-work/p21-framework](http://www.p21.org/our-work/p21-framework)

- Peters-Burton, E. E., Merz, S. A., Ramirez, E. M., & Saroughi, M. (2015). The effect of cognitive apprenticeship-based professional development on teacher self-efficacy of science teaching, motivation, knowledge calibration, and perceptions of inquiry-based teaching. *Journal of Science Teacher Education*, 26(6), 525–548. <https://doi.org/10.1007/s10972-015-9436-1>
- Ponjuan, L., McKinney, L., Waltrip, L., & Sadler, T. D. (2007). *Engaging future scientists: Assessing undergraduate research experiences for students of color*. Louisville, KY: Association for the Study of Higher Education.
- Pop, M. M., Dixon, P., & Grove, C. M. (2010). Research experiences for teachers (RET): Motivation, expectations, and changes to teaching practices due to professional program involvement. *Journal of Science Teacher Education*, 21(2), 127–147. <https://doi.org/10.1007/s10972-009-9167-2>
- 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 August 5, 2019, from <https://peer.asee.org/27710>
- Remich, R., Naffziger-Hirsch, M. E., Gazley, J. L., & McGee, R. (2016). Scientific growth and identity development during a postbaccalaureate program: Results from a multisite qualitative study. *CBE—Life Sciences Education*, 15(3), ar25. <https://doi.org/10.1187/cbe.16-01-0035>
- Robnett, R. D., Chemers, M. M., & Zurbruggen, E. L. (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. *Journal of Research in Science Teaching*, 52(6), 847–867. <https://doi.org/10.1002/tea.21221>
- Roehrig, G. H., Moore, T. J., Wang, H. H., & Park, M. S. (2012). Is adding the E enough? Investigating the impact of K–12 engineering standards on the implementation of STEM integration. *School Science and Mathematics*, 112(1), 31–44. <https://doi.org/10.1111/j.1949-8594.2011.00112.x>
- Roth, W. M. (1995). *Authentic school science: Knowing and learning in open-inquiry science laboratories*. Dordrecht, Netherlands: Kluwer Academic.
- Sadler, T. D., Burgin, S., McKinney, L., & Ponjuan, L. (2010). Learning science through research apprenticeships: A critical review of the literature. *Journal of Research in Science Teaching*, 47(3), 235–256. <https://doi.org/10.1002/tea.20326>
- Sadler, T. D., & McKinney, L. (2010). Scientific research for undergraduate students: A review of the literature. *Journal of College Science Teaching*, 39(5), 43–49.
- Saka, Y. (2013). Who are the science teachers that seek professional development in research experience for teachers (RET's)? Implications for teacher professional development. *Journal of Science Education and Technology*, 22(6), 934–951. <https://doi.org/10.1007/s10956-013-9440-1>
- Salto, L. M., Riggs, M. L., De Leon, D. D., Casiano, C. A., & De Leon, M. (2014). Underrepresented minority high school and college students report STEM-pipeline sustaining gains after participating in the Loma Linda University summer health disparities research program. *PLoS ONE*, 9(9), 1–13. <https://doi.org/10.1371/journal.pone.0108497>
- Sampson, V., Enderle, P. J., & Walker, J. P. (2011). The development and validation of the Assessment of Scientific Argumentation in the Classroom (ASAC) observation protocol: A tool for evaluating how students participate in scientific argumentation. In Khine, M. (Ed.), *Perspectives on scientific argumentation: Theory, practice and research* (pp. 235–264). New York: Springer.
- Schwartz, J. J. (2012). Faculty as undergraduate research mentors for students of color: Taking into account the costs. *Science Education*, 96(3), 527–542. <https://doi.org/10.1002/sce.21004>
- Schwartz, R., Westerlund, J., Garcia, D., & Taylor, T. (2010). The impact of full immersion scientific research experiences on teachers' views of nature of science. *Electronic Journal of Science Education*, 14(1), 1–40.
- Schwartz, R. S., & Crawford, B. A. (2004). Authentic scientific inquiry as a context for teaching nature of science: Identifying critical elements for success. In Flick L. & Lederman, N. (Eds.), *Scientific inquiry and nature of science: Implications for teaching, learning, and teacher education* (pp. 331–355). Dordrecht, Netherlands: Kluwer Academic.
- Schwartz, R. S., Lederman, N. G., & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Education*, 88(4), 610–645. <https://doi.org/10.1002/sce.10128>
- Shaffer, C. D., Alvarez, C. J., Bednarski, A. E., Dunbar, D., Goodman, A. L., Reinke, C., ... & Bazinet, C. (2014). A course-based research experience: How benefits change with increased investment in instructional time. *CBE—Life Sciences Education*, 13(1), 111–130.
- Shapiro, C., Moberg-Parker, J., Toma, S., Ayon, C., Zimmerman, H., Roth-Johnson, E. A., ... & Sanders, E. R. (2015). Comparing the impact of course-based and apprentice-based research experiences in a life science laboratory curriculum. *Journal of Microbiology & Biology Education*, 16(2), 186–197. <https://doi.org/10.1128/jmbe.v16i2.1045>
- Silverstein, S. C., Dubner, J., Miller, J., Glied, S., & Loike, J. D. (2009). Teachers' participation in research programs improves their students' achievement in science. *Science*, 326(5951), 440–442. <https://doi.org/10.1126/science.1177344>
- Slovacek, S., Whittinghill, J., Flenoury, & Wiseman, D. (2012). Promoting minority success in the sciences: The minority opportunities in research programs at CSULA. *Journal of Research in Science Teaching*, 49(2), 199–217. <https://doi.org/10.1002/tea.20451>
- Southerland, S. A., Granger, E. M., Hughes, R., Enderle, P., Ke, F., Roseler, K., ... & Tekkumru-Kisa, M. (2016). Essential aspects of science teacher professional development: Making research participation instructionally effective. *AERA Open*, 2(4), 1–16. <https://doi.org/10.1177/2332858416674200>
- Staub, N. L., Poxleitner, M., Braley, A., Smith-Flores, H., Pribbenow, C. M., Jaworski, L., ... & Anders, K. R. (2016). Scaling up: Adapting a phage-hunting course to increase participation of first-year students in research. *CBE—Life Sciences Education*, 15(2), ar13. <https://doi.org/10.1187/cbe.15-10-0211>
- Stephens, A. L., Pallant, A., & McIntyre, C. (2016). Telepresence-enabled remote fieldwork: Undergraduate research in the deep sea. *International Journal of Science Education*, 38(13), 2096–2113. <https://doi.org/10.1080/09500693.2016.1228128>
- Stevens, L. M., & Hoskins, S. G. (2014). The CREATE strategy for intensive analysis of primary literature can be used effectively by newly trained faculty to produce multiple gains in diverse students. *CBE—Life Sciences Education*, 13(2), 224–242. <https://doi.org/10.1187/cbe.13-12-0239>
- Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, 40(3), 371–384. <https://doi.org/10.2307/2393788>
- Thompson, J. J., Conaway, E., & Dolan, E. L. (2015). Undergraduate students' development of social, cultural, and human capital in a networked research experience. *Cultural Studies of Science Education*, 11(4), 959–990. <https://doi.org/10.1007/s11422-014-9628-6>
- Tootle, T. L., Hoffmann, D. S., Allen, A. K., Spracklen, A. J., Groen, C. M., & Kelsch, D. J. (2019). Mini-course-based undergraduate research experience: Impact on student understanding of STEM research and interest in STEM programs. *Journal of College Science Teaching*, 48(6), 44–54.
- University of Arizona. (2018). *Teachers in industry*. Retrieved December 1, 2018, from <https://teachersinindustry.arizona.edu/>
- Virginia Science and Technology Campus. (2018). *GW Teachers in Industry Project*. Retrieved December 14, 2018, from <https://virginia.gwu.edu/teachers-industry>
- Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems Thinker*, 9(5), 1–10.
- Yaffe, K., Bender, C., & Sechrest, L. (2014). How does undergraduate research experience impact career trajectories and level of career satisfaction: A comparative survey. *Journal of College Science Teaching*, 44(1), 25–33.
- Yelamarthi, K., Kaya, T., DeJong, B., Chen, D., Hu, Q., & Cheng, F. (2013). An engineering research program for high school science teachers: Feedback and lessons learned from the pilot implementation. *Technology Interface International Journal*, 13(2), 49–60.
- Zhou, Y., Jung, E., Arroyave, R., Radovic, M., & Shamberger, P. (2015). Incorporating research experiences into an introductory materials science course. *International Journal of Engineering Education*, 31(6), 1491–1503.