AAAS | ARISE

Advancing Research & Innovation
in the STEM Education of
Preservice Teachers in High-Need School Districts

A Synthesis of Research on and Measurement of STEM Teacher Preparation

Courtney A. Bell
Drew H. Gitomer
Corey Savage
Anneliese Haines McKenna

A 2019 AAAS ARISE Commissioned Paper
Prepared for the
American Association for the Advancement of Science

A Synthesis of Research on and Measurement of STEM Teacher Preparation

Part of the 2019 AAAS ARISE Commissioned Paper Series
Prepared for the American Association for the Advancement of Science

Courtney A. Bell
Educational Testing Service

Drew H. Gitomer
Rutgers University

Corey Savage
University of Tübingen

Anneliese Haines McKenna
Rutgers University

Contents

Acknowledgements

This research was supported by the National Science Foundation (NSF) under Grant No. DUE-1548986 to the American Association for the Advancement of Science (AAAS). Any opinions, findings, interpretations, conclusions or recommendations expressed in this material are those of its authors and do not represent the views of the AAAS Board of Directors, the Council of AAAS, AAAS' membership or the National Science Foundation.

2019 AAAS ARISE Commissioned Paper Series

Wilson, S. M. (2019). *Introduction to 2019 AAAS ARISE Commissioned Paper Series*. American Association for the Advancement of Science.

1. Bell, C., Gitomer, D., Savage, C., & Mckenna, A. H. (2019). *A Synthesis of Research on and Measurement of STEM Teacher Preparation*. American Association for the Advancement of Science.

2. Fuller, E. J., & Pendola, A. (2019). *Teacher Preparation and Teacher Retention: Examining the Relationship for Beginning STEM Teachers*. American Association for the Advancement of Science.

3. Youngs, P., Bieda, K., & Kim, J. (2019). *Teacher Induction Programs Associated with Retention in the STEM Teaching Workforce*. American Association for the Advancement of Science.

Read the full papers at: https://aaas-arise.org/commissioned-papers/.

Suggested Citation

Bell, C., Gitomer, D., Savage, C., & Mckenna, A. H. (2019). *A Synthesis of Research on and Measurement of STEM Teacher Preparation*. American Association for the Advancement of Science.

Abstract

There are research syntheses that review what the field knows about various aspects of STEM teacher preparation (e.g., National Research Council, 2000; Wilson, 2011) and reviews of teacher preparation across subjects (e.g., Cochran-Smith & Zeichner, 2013; Cochran-Smith, Villegas, Abrams, Chavez-Moreno, Mills & Stern, 2016). This review takes up yet another, related topic -- the research and measurement approaches used to study STEM teacher preparation. Drawing on recent articles in STEM and general teacher education journals, the review takes a situated perspective and categorizes the research into seven inductively developed purposes: understanding STEM preservice teacher learning and development, improving educator preparation programs, contributing to program accountability, describing and understanding relationships between STEM preparation and valued outcomes, understanding assessments and measurement of STEM preparation, framing and reframing issues of STEM preparation, and understanding teacher educators and their practices. Within each of these purposes, the review summarizes the questions and phenomena under investigation and the methodological approaches used to understand these questions and phenomena. The authors offer insights about the questions and phenomena that have not yet been addressed in each purpose and suggest varied research agendas that could help the field strengthen research on and measurement of STEM teacher preparation.

Introduction

What is the status of research on and measurement of Science, Technology, Engineering, and Mathematics (STEM) teacher preparation? While handbooks and research articles (e.g., (National Research Council [NRC], 2000; Wilson, 2011) regularly summarize what we know about specific phenomena within STEM teacher preparation, it is less common to review the research and measurement approaches taken to study STEM teacher preparation. That is the focus of this review.

The topic is large enough that many chapters could be written about research on and measurement of STEM teacher preparation. The issues attendant to such study are varied; for example: What types of questions are asked by researchers? What instrumentation is used? How has research changed over time? Given the charge of this paper, we bound the synthesis to consider the following: the kinds of questions that are addressed; the methodologies and measures used to produce research evidence; the characteristics of the findings produced by the field; and suggestions for improving the quality and impact of this broad research agenda. We limited our search to the most recent three years of research activity for two related reasons. First, a preliminary reading of earlier reviews and commentaries of research (e.g., Arbaugh, Ball, Grossman, Heller, & Monk, 2015; Cochran-Smith & Villegas, 2016) made clear that there have not been significant shifts in the methodologies used to study teacher preparation. Second, extending the timeframe of the review would have increased the length of the paper without altering its findings substantially.

It is necessary to clarify three uses of terminology that are used throughout this paper. First, what is included in *STEM teacher preparation*? For purposes of this review, the term, *STEM teacher preparation*, refers to the teacher candidate's context and experience, including learning that occurs prior to the pre-service teacher (PST) beginning to teach for the first time as the teacher of record in a STEM subject in U.S. public schools. This includes the study of individuals who are learning to teach STEM as well as the people and institutions engaged in preparing those teachers. Those institutions may be university-based and/or non-university-based, “traditional” or “alternative.” We do not address in-service professional development efforts within this review.

Second, the term, *STEM teacher preparation*, refers to teachers spanning the entire K–12 range. Kindergarten through the middle grades is a formative time for K–12 students learning STEM subjects (e.g., Ma, 1999).[1] This means that we must consider the preparation of elementary school teachers to be part of STEM teacher preparation even though many elementary teachers do not feel well qualified to teach science (Trygstad, Smith, Banilower, & Nelson, 2013). Further, some teachers become certified to teach and begin teaching in a non-STEM field and then later go through additional preparation activities to become certified in a STEM field. These teachers and programs are also included in our definition of *STEM teacher preparation*.

Third, we distinguish the following terms: *research, methods, methodologies,* and *measures*. *Research* is an inquiry activity carried out by a range of stakeholders—policymakers, researchers, practitioners—for various purposes. Although there are many definitions of *research*, we adopt Creswell’s (2012) general assertion that research “is a process of steps used to collect and analyze information to increase our understanding of a topic or issue” (p. 4).

Methods are the “tools that researchers use to do their work” (NRC, 2005) in order to make sense of social phenomena and include, for example, questionnaires, interviews, observations, regressions, and experiments. *Methodologies* encompass multiple methods and reflect the basic principles and processes used to carry out research (Moss & Haertel, 2016).

Methodologies are defined by procedures through which research is designed and enacted; data is collected or produced; analytic approaches are conducted; and interpretations and evaluations are made (e.g., Moss & Haertel, 2016). The appropriateness of any methodology can only be judged in the context of the questions they are being used to answer (e.g., NRC, 2002). Methodologies are designed to address particular types of research questions and to provide certain kinds of insights. No single methodology can address all of the important questions about a given phenomenon.[2] Thus, it is important that multiple methodologies are represented in answering questions about STEM teacher preparation both within and across research studies.

Finally, there are *measures*, which are the protocols, tools, or criteria that are employed for collecting data when enacting the methods. Measures are often used to describe tools or instruments that result in scaled quantitative distinctions such as those associated with assessments or observation protocols. For purposes of this report, we take a broader perspective and use *measure* to refer to any systematic protocol used to collect, analyze, and represent data. For example, while surveys represent a particular type of research method, there are specific survey measures intended to capture evidence about some particular construct or set of constructs. Thus, when we validate or evaluate a measure or instrument, we only consider that measure, not the entire method. In this review, we focus primarily on methods that are used to support research on STEM teacher preparation, including interviews, observations, and document analysis as well as surveys and more traditionally considered standardized assessments (i.e., tests of knowledge, portfolios, etc.), and do not delve into discussions of particular measures.

Given these definitions, judgments of research activities should be placed in their social and cultural contexts. Research studies can focus on individual educator preparation programs (EPPs) and their pre-service teachers (PSTs) (e.g., a study of how pre-service teachers learn to teach diverse students), or they can address issues that examine much broader patterns (e.g., a national impact study of a policy requiring that all science teachers have a subject-matter major). The measures used in such research can be formal instruments (e.g., a portfolio-based performance assessment of a pre-service teacher's learning over her preparation program or a two-hour standardized multiple-choice test of a candidate's knowledge of mathematics or a process) or more informal measures such as an annual review of the quality and appropriateness of the mentor teachers with whom student teachers are placed.

This report is organized in terms of seven purposes of research on STEM teacher preparation and the methods and measures used to support this research. As we read through and organized the research, we inductively developed and refined these seven categories, as the scholarship reviewed for this report did not always explicitly name purposes in the same terms as presented in this paper. While some overlap exists across purposes, studies tended to focus on one of these seven broad research purposes:

> *Understanding STEM PST learning and development*: These studies are primarily concerned with documenting the learning and development process of PSTs; examining PST learning outcomes in the areas of knowledge, beliefs, and/or practices; or studying interventions at the course, classroom, or instructor level.

> *Improving EPPs*: These studies examine or evaluate interventions that are intended to improve programs. The focus of these studies is at the program level, in contrast to the previous purpose, which is focused on experiences, courses, classrooms, or instructors.

Contributing to EPP accountability: These studies add evidence and methodological insights to research and policy debates focused on accountability issues in STEM PST teacher preparation.

Describing and understanding relationships between STEM teacher preparation and other valued outcomes: These studies focus on the relationships between STEM teacher preparation and practicing teacher and student processes and outcomes, including but not limited to student achievement and teacher labor market decisions.

Understanding assessments and measurement of STEM teacher preparation quality: These studies focus on the validity and reliability of measures of STEM teacher preparation quality including value-added estimates and teacher performance assessments.

Framing or reframing issues of STEM teacher preparation: These studies include discussions of historical trends, reviews of literature, and constructions and/or critiques of conceptual frameworks, all used to consider different dimensions of STEM teacher preparation.

Understanding teacher educators and their practices: These studies focus on STEM teacher educators as individuals as well as how they learn to carry out the work of teaching PSTs. Research examines teacher educators' knowledge, practices, capabilities, beliefs, and identities as characteristics important to the learning opportunities provided to PSTs.

In this paper, our aim is to summarize the nature of research activity addressing each purpose and then make recommendations for future research directions, given the current state of research and methodology within those purposes. In order to do this, for each purpose we address the following questions:

1. What are the most important questions and phenomena studied over the past three years within this purpose?
2. What methodological approaches have been used to understand these questions and phenomena?
3. Given the range of issues associated with this purpose, what are some of the most important questions and phenomena that have *not* been addressed?
4. What type of research agenda might be productively engaged moving forward?

Analytic Approach

Given the scope of this paper, a comprehensive review going back a large number of years was not feasible. Instead, our analysis is built on two review activities—a systematic, bounded review of the research and a non-systematic gathering of relevant research and assessment documents that pertain to the purpose of our review. The systematic review began with a search of 14 peer-reviewed journals published from January 1, 2014–September 30, 2017. The authors identified the journals to sample across types of journals that publish relevant research, including leading journals published in the United States and internationally ranging from general educational research to those focused specifically on STEM and/or teacher preparation. The number of articles reviewed that met our search criteria and their associated journals is presented in Table 1; the listing of those articles is in the Appendix.

The authors used the following terms to search the selected journals: *STEM Education, Science Education, Mathematics Education, Engineering Education, Teacher Education, Teacher Preparation, Educator Preparation, Teacher Training*, and *Alternative Certification*.

The search yielded 1,466 unique results. After several rounds of abstract reading we discarded journal articles that were not about pre-service teacher preparation; that were not U. S.-focused or international with a mention of the United States; and that were not elementary or Grades 6–12 STEM-focused. We also discarded journal articles that were not considered research in and of themselves (e.g., book reviews, introductions to journal issues, opinion pieces). Following several rounds of study selection, 174 unique articles (approximately 12%) were included for a full review. These 174 articles were then uploaded into Endnote. Finally, we read and summarized each article in terms of research goals, researcher's role, sample, methods, research question(s), phenomena of inquiry, findings, and validity. Table 2 presents a more detailed description of the summarization categories.

During this process any inconsistencies between readers were discussed at a weekly meeting. After the in-depth reading of these 174 articles, we sorted the articles by the previously articulated purposes implicit or explicit in each article. Three articles were assigned to two purposes. This sorting led to the following results: Understanding STEM PST learning and development (104 articles); Improving EPPs (1 article); Contributing to EPP accountability (2 articles); Describing and understanding relationships between STEM teacher preparation and other valued outcomes (22 articles); Understanding measures of STEM teacher preparation quality (5 articles); Framing or reframing issues of STEM teacher preparation (29 articles); and Understanding teacher educators and their practice (14 articles). The authors then reviewed each group of articles for the four research questions developed for each purpose, noting themes and patterns. Discrepancies were discussed at regular meetings.

A significant proportion of articles that met our search criteria did not have clear implications for STEM preparation and/or were not focused exclusively on STEM populations.

For example, an article might have been an empirical report about elementary pre-service teachers learning how to teach students in an urban context, but the findings were written in a general format and did not draw specific implications for STEM teachers. Such an article certainly offers relevant insights and, in the interest of being inclusive, we did not discard it. However, the directness of insights from such an article is somewhat different than the same article focused only on STEM elementary teachers. Thus, for each of the 174 articles we asked two questions: 1) Is the population of study participants exclusively STEM teachers? and 2) Are the findings framed such that there are direct implications for STEM teacher preparation? Articles were considered as "focusing on STEM" as long as the answer to at least one of the questions was yes. For 66 articles the answer to both questions was no—though we include these articles in our review we note in the Appendix their status as non-exclusively STEM.

Our non-systematic review produced many relevant documents about STEM teacher preparation that are not found in journals. These are often reports that represent research summaries as well as policy perspectives built on research. It also produced journal articles that fell outside of the systematic review either due to journal or date. We reference these reports and articles when relevant.

It is important to note that there is a body of literature in the economics of education that has examined pre-service teacher education but is not included in our review. They are not included, first, because we did not include any discipline-specific journals (e.g., sociology, economics, etc.). Further, while some studies have examined how aspects of teacher preparation or preparation programs relate to outcomes such as K–12 students' mathematics and language arts test scores, they generally do not focus explicitly on STEM teacher preparation nor have they looked explicitly at STEM K–12 placements. Even when the studies do include STEM programs, researchers often do not report results in a disaggregated fashion such that we could isolate STEM teacher preparation findings. These studies, therefore, contribute to an understanding of teacher preparation broadly conceived but do little to help us understand STEM preparation. Despite this, such studies nominate and describe important relationships between variables. To the degree these articles appeared in our non-systematic search, we note them. We also discuss the larger issue of how we might learn more about STEM teacher preparation from such studies.

Criteria to Judge Research on and Assessment of STEM Teacher Preparation

There is not a single set of criteria against which we can judge research on and measurement of STEM teacher preparation. Criteria for judging research can be applied at more micro- or macro-levels. A *micro-level* approach considers the quality of individual research studies. Such judgments might consider the quality of the claims or assess the specific sampling or analysis approach taken by researchers. We used peer-review journals of some reputation to serve as a proxy for this type of micro-review. Our goal is to understand the broad trends in research on STEM teacher preparation and, therefore, take a more *macro-level* perspective. That perspective focuses across studies on the general characteristics of the studies (e.g., the focus of the research questions, the methodologies used, the nature and size of the samples) within one of the seven purposes.

Description of Summarization Categories

An original set of categories was developed based on our understanding of previous reviews of teacher preparation research (e.g., Cochran-Smith & Villegas, 2016; NRC; 2010; Wilson, Floden, & Ferrini-Mundy, 2001) and the ways in which that research varies; however, those categories were revised as the team began to summarize articles. Revisions reflect problems in our original definitions of categories and variation that we did not consider at the outset. We summarize the research along with the final set of summarization categories in Table 2.

Broad categories. The first category refers to the seven previously described research purposes for studying STEM teacher preparation. The next category refers to the role of the researcher and the relationship between the researcher and the parties under examination. Researchers may study their own contexts of practice, or they may have no personal connections to those who are researched; both stances (and the stances in-between) influence how studies are conducted and interpreted (see Moss & Haertel, 2016). This is especially noted in reviews of teacher preparation (e.g., Cochran-Smith & Villegas, 2016).

We then characterized the nature of the study sample, considering study design with respect to grade level, subject-matter focus, and participant race/ethnicity. We also examined sample size and the nature of the sampling approach.

Each study was then categorized in terms of its methodological approach. While this paper cannot adequately represent any of these methods or methodologies, there are certain core aspects that are highlighted as central to the study of STEM teacher preparation and the claims that can be made on the basis of such research. We briefly describe the most commonly used methodologies in STEM teacher preparation research. For a much fuller treatment of methodologies used to study teaching, see Moss and Haertel (2016).

Common methods in teacher preparation research. A critical question in understanding teacher preparation is whether some type of intervention actually causes some change in a valued outcome such as learning by PSTs or teacher educators. The ideal research design to establish the causal relationship of a variable to an outcome is to conduct an experiment in which treatment conditions are randomly assigned to the unit of analysis (typically PSTs or preparation institutions). However, in most cases, it is not possible to carry out true experiments.

Therefore, researchers turn to a range of quasi-experimental methodologies (Shadish, Cook, & Campbell, 2001) that attempt to approximate random assignment by controlling for other factors aside from the variable(s) of interest. A common quasi-experimental method used in studies of teaching generally involves the use of covariates within regression analyses using large-scale data sets. The goal of such analyses is to estimate the effect of a particular treatment variable on some outcome while controlling for as many variables as possible. While such approaches have been widely used to justify educational policies, their ability to establish causal relationships is questionable. Moss and Haertel (2016) note that, "[b]y themselves, covariate adjustments cannot offer a rigorous warrant for inferences as to the particular causes for observed differences among nonequivalent groups" (p. 138). Importantly, both experiments and quasi-experiments, if carried out without additional types of methodological approaches, only allow researchers to estimate the presence and magnitude of an effect. They do not provide insight into why or how such an effect is produced. These studies, however, can begin to establish an empirical database of findings that can be evaluated across studies. For research questions where study results converge, there is more confidence to support particular policy decisions.

Other common approaches to examining relationships in teacher preparation involve self-report methodologies using survey and interview methods. For surveys, all participants are given the same questions; response options are typically highly constrained and developed by the researchers; and responses are summarized using quantitative techniques. Surveys often seek to measure multiple constructs (e.g., self-efficacy, satisfaction with one's preparation program, beliefs about teaching); these constructs are operationalized in scales. Interviews can allow study participants to explain their views in their own categories and words. Interviews can range from highly unstructured and open-ended to semi-structured in which exchanges begin with common questions for all respondents but then can involve more unique follow-ups and exchanges between interviewer and respondent.

Critical methodologies, particularly those that build on Critical Race Theory (CRT), have been used to understand schools and teaching, including teacher preparation. The basic premise is to use race (or ethnicity, sexuality, etc.) as an analytic tool to understand societal inequity (Ladson-Billings & Tate, 1995). Such approaches are built on the understanding that race is a significant factor in determining educational inequity, that research and society themselves are racialized, and that all research is done in some group's self-interest.

There are other qualitative approaches that attempt to understand particular contexts or cases. Ethnographic methodologies frequently are characterized by extended participant observation in a single site (e.g., teacher preparation institution) (e.g., Eisenhart, 2001; Erickson, 1986; Moss & Haertel, 2016) and the use of anthropological research methods to make interpretive judgments about the interactions and artifacts that occur within the particular context.

Case study methodologies either focus on a single case (e.g., a program, a course, an individual) or can involve comparisons of a small number of cases. In contrast to more quantitative approaches, the focus of case studies is on the case, not on a particular set of variables (e.g., George & Bennett, 2004; Moss & Haertel, 2016). Researchers attend to conditions and mechanisms that are associated with particular outcomes. Case studies can involve a range of methods (e.g., surveys, interviews, observation, critical perspectives) and methodologies.

For methodologies that fundamentally rely on interpretations of discourse, observations, interactions, artifacts, documents, policies, as well as quantitative data (e.g., proportion of teachers of color in STEM teacher preparation), study findings will be bound to the study's contexts in ways that generally do not allow for statistical generalization beyond the contexts studied. Instead, the goal of generalization in these studies focuses on theory—to develop theoretical understandings of particular contexts that can support researchers and users of research in making sense of the same or related phenomena in other contexts.

Validity of measures. Across methodological approaches, the validity of findings is, in large part, dependent on the validity of the measures used. There is an agreed-upon set of standards (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) that guides appraisals of measurement and measures—many of these standards applying directly to measures used by STEM teacher preparation programs (e.g., a field supervisor's use of an observation rubric to assess a student teacher's performance) as well as assessments of teacher preparation quality (e.g., program-wide portfolios of teaching competency, value-added measure [VAM] scores for EPPs). We identify three basic criteria that all measures should meet and apply those in our review: accuracy, reliability, and the developmental stage of the measure.

Scores should be accurate. When measures are accurate, they reflect the actual level of knowledge or skill the person has at that time. There are many ways in which scores from measures may be inaccurate. In STEM teacher preparation research, score inaccuracies may commonly result from using human raters (or observers) to create ratings of PSTs' teaching practices or written essays. In assessments that use human raters to create scores, accuracy should receive additional attention (AERA et al., 2014, Standards 6.8, 6.9, pp. 112, 118). Human raters are known to be somewhat inaccurate when observing videos of teaching (e.g., Bill and Melinda Gates Foundation, 2012), and after they have been trained, they frequently do not use the cognitive processes they are trained to use (e.g., Bell et al., 2013). They also tend to drift and are not stable in their severity over time (e.g., Casabianca, Lockwood, & McCaffrey, 2014). These realities suggest that accuracy is particularly important when measures of preparation include human raters.

Scores should be reliable. Scores should also be consistent. In technical terms, this means the scores are not systematically influenced by factors unrelated to the construct being measured. For STEM teacher preparation, the construct measured is often teaching practice, teacher knowledge, or teacher quality. If a candidate takes a knowledge test with particular items (called a test form) as a requirement for licensure, it should not matter if she took form one or form two. The tests should be internally reliable enough that if the candidate has a certain level of knowledge, the test scores from different groups of items or forms will reliably reflect this. In performance assessments such as observations of student teachers or portfolios used at the end of a program to document a candidate's proficiency in specific teaching practices, scores should be sufficiently reliable that it does not matter who scored the assessments, when the assessments were completed, or what specific subject-matter and grade-level combinations were assessed. Scales on assessments and surveys, observation scores, and other measures created by raters all have various reliabilities associated with them. These reliability metrics should be reported.

Measures should develop validity arguments over time. Reliability and accuracy are foundational to any measure's validity; however, when a measure is used there should be some evidence that the measure's scores mean what researchers think they mean. To ascertain this, researchers must collect additional evidence about the measure and develop an argument for the appropriateness of the score use (Kane, 2006). That argument is a combination of logical and empirical evidence that is developed over multiple studies. There is no single study that alone can establish the validity of a particular measurement instrument. Instead, multiple studies carried out over time by different researchers contribute to the field's understanding of a specific measure's validity argument. Teacher licensure exams are a good example of this. Various researchers have studied different aspects of the assessments including what they measure and their reliability (NRC, 2000), predictive validity (Goldhaber, 2007), and relationships with other known measures (Clotfelter, Ladd, & Vigdor, 2007). The study of teacher preparation requires some measures that are local and specific to the context being studied, but over time the field should also develop measures that have elaborated validity arguments, implying that at least some research studies should use the same or very similar measures over time.

The Focus of Teacher Preparation Research Questions

Researchers focus on specific objects of inquiry. In addition to focusing on the actual participants being studied—pre-service teachers, teacher educators, educator preparation programs—researchers focus on important constructs. The diversity of foci is large; however, we can group these constructs broadly into knowledge, practices, and attitudes and beliefs.

Teacher knowledge. That teachers should have an agreed-upon body of knowledge has been a given since the first large-scale standardized teacher tests were introduced in 1940. The specifics of what should be known, however, has evolved during the intervening 80 years and now is focused on the development of content knowledge (CK), pedagogical content knowledge (PCK), and/or content knowledge for teaching (CKT; Gitomer & Zisk, 2015). Common measures of teacher knowledge that have a base of validity evidence include licensure tests such as the *Praxis*® (see https://www.ets.org/praxis/faq_test_takers/) as well as measures of CKT such as *Mathematical Knowledge for Teaching* (Hill, Rowan, & Ball, 2005; Kersting, Givvin, Thompson, Santagata, & Stigler, 2012).

Teacher practice. There are now broad theories that build on teaching research about the kinds of practices that characterize effective teaching (Ball & Forzani, 2009; Grossman & McDonald, 2008). Many of these practices are measured through general observation (e.g., *Classroom Assessment Scoring System* [CLASS™]; Pianta, Hamre, Haynes, Mintz, & La Paro, 2005; *Framework for Teaching* [FFT]; Danielson, 2011) as well as STEM-specific protocols (*Mathematical Quality of Instruction* [MQI]; Hill et al., 2008; *Inside the Classroom: Observation and Analytic Protocol;* Horizon Research, Inc., 2000). These protocols have been widely used in research on teaching, and a body of validity evidence exists for each. Other practices, particularly those associated with the quality of instructional artifacts, including assessments and assignments, have also been studied with common research instruments in mathematics and science (*Intellectual Demand Assignment Protocol* [IDAP]; Wenzel, Nagaoka, Morris, Billings, & Fendt, 2002; QAS; Martínez, Borko, Stecher, Luskin, & Kloser, 2012; SCOOP; Borko, Stecher, & Kuffner, 2007), and validation evidence has been produced.

Attitudes and beliefs. A wide range of attitudes and beliefs has been studied—for example, beliefs about content such as the nature of science or evolution (Andrà, Brunetto, Levenson, & Liljedahl, 2017; Katsh-Singer, McNeill, & Loper, 2016) as well as beliefs about teaching, particularly self-efficacy (Tschannen-Moran, Hoy, & Hoy, 1998). Attitudes and beliefs play important roles in both knowledge and practices and, therefore, are another critical component of teacher preparation activities.

In the following sections we consider each of the seven purposes in terms of the status of research on pre-service teacher preparation.

Understanding STEM PST Learning and Development

Of the 174 articles reviewed, more than half of them, 104, focused on documenting the nature of PSTs' learning and development. All but seven of the articles satisfied the search criteria for being STEM-focused.

What Are the Important Questions and Phenomena Being Addressed?

The vast majority of studies explores the development of PST knowledge, practices, and/or attitudes and beliefs in the context of a program or a particular pedagogical or curricular intervention. While the studies are largely descriptive, they generally investigate questions of how particular features of STEM teacher preparation affect PSTs' knowledge, practice, and/or attitudes and beliefs, as well as their ability to reflect on aspects of their own practice or to analyze the practice of others. Major foci of studies have included dimensions of teaching quality that have been identified through research (e.g., Cochran-Smith, Villegas et al., 2016) and in policy (Council for the Accreditation of Educator Preparation [CAEP], 2016a).

Teacher knowledge. Some studies have simply tried to understand the structure of PSTs' knowledge about important concepts they will teach. For example, Lovin, Stevens, Siegfried, Wilkins, and Norton (2018) probed how PreK–8 teachers understand mathematical fractions (CK) while Lannin et al. (2013) investigated teachers' development of PCK over the course of the field placement and first year of teaching mathematics. Other studies have explored how knowledge develops within STEM teacher preparation programs, often attributing development to particular features of the program. For example, mathematics education researchers (Hohensee, 2017; Reeves & Honig, 2015; Thanheiser, 2015; Whitacre & Nickerson, 2016) have examined how PSTs develop understandings of mathematics concepts through coursework or through other teacher education curricula. Parallel studies in the science domain have been carried out (e.g., Akerson, Khemmawadee, Park Rogers, Carter, & Galindo, 2017; Donna & Hick, 2017; Johnson & Cotterman, 2015; Saçkes & Trundle, 2014; Santau, Maerten-Rivera, Bovis, & Orend, 2014; Thanheiser, 2015). Other studies (e.g., Barnett & Friedrichsen, 2015; Diezmann & Watters, 2015) have explored such issues as how pre-service education can support PCK development for PSTs.

Teacher practice. Research has focused on the development of specific STEM teaching practices within EPPs. A number of studies (e.g., Benedict-Chambers & Aram, 2017; Kang & Anderson, 2015; Mitchell & Marin, 2015; Roller, 2016; Sun & van Es, 2015; Weiland, Hudson, & Amador, 2014) have focused on teacher noticing directed toward either student thinking and learning or their own practice. Others have focused on teacher education efforts to improve core practices of STEM teaching (e.g., Ball & Forzani, 2009; Grossman & McDonald, 2008) such as holding classroom discussion (Ghousseini, 2015; Ghousseini & Herbst, 2016; Tyminski, Zambak, Drake, & Land, 2014); engaging in scientific inquiry (e.g., McNew-Birren & van den Kieboom, 2017); conducting formative assessment (e.g., Sabel, Forbes, & Zangori, 2015; Santagata & Yeh, 2014; Weiland et al., 2014); building and using representations (Ghousseini & Herbst, 2016; Hohensee, 2017; Subramaniam, 2014); and developing culturally responsive teaching practices and understandings (Bottoms, Ciechanowski, Jones, de la Hoz, & Fonseca, 2017; Rubin, Abu El-Haj, Graham, & Clay, 2016).

A final teacher practice research focus has been PST teachers' ability to reflect on their practice and/or professional learning (e.g., Moore-Russo & Wilsey, 2014; Ronfeldt & Reininger, 2012; Ronfeldt, Reininger, & Kwok, 2013; Saçkes & Trundle, 2014; Santagata & Yeh, 2014; Thomson & Palermo, 2014) or to analyze the teaching of others (Bravo, Mosqueda, Solís, & Stoddart, 2014; Olson, Bruxvoort, & Vande Haar, 2016; Yeh & Santagata, 2015).

Attitudes and beliefs. A number of studies have explored PSTs' attitudes and beliefs toward STEM teaching, often studying how attitudes and beliefs change over the course of TE or as a result of a particular intervention. Some studies have explored teachers' sense of self-efficacy (Tschannen-Moran et al., 1998) in teaching particular subjects or content areas (Bautista & Boone, 2015; Menon & Sadler, 2016, 2018). Some studies have focused on teachers' beliefs about STEM content and the teaching of that content (Jacobson, 2017; Jong & Hodges, 2015; Reeves & Honig, 2015; Riegle-Crumb et al., 2015). Others have examined beliefs and belief changes related to teaching linguistically, ethnically, racially, and economically diverse student populations (Bravo et al., 2014; Kesner, Kwon, & Lim, 2016; Pappamihiel, Ousley-Exum, & Ritzhaupt, 2017; Reagan, Chen, & Vernikoff, 2016). Still other studies have examined how teacher beliefs change during TE with respect to the use of technology tools to support STEM teaching and learning (Beilstein, Perry, & Bates, 2017; Brown, Englehardt, & Mathers, 2016).

The Sample: Who is Being Studied, and Who is Doing the Research?

Of the studies included in the formal review, almost two-fifths focused on mathematics, and a similar proportion focused on science. None focused on engineering or technology. Almost 5% examined mathematics and science preparation, and the remaining studies considered mathematics and science as part of larger efforts that included non-STEM subjects as well. Approximately half of the studies focused on elementary preparation, and one quarter examined secondary teacher preparation. The remaining studies considered either all grade levels or early childhood while a handful of studies were not explicit in their grade-level focus. Only 38% of the studies reported on the racial background of the teacher participants. Of these 40 studies, 30 had predominantly White participants (i.e., more than 80% of the study's sample reported that they were White). And of the remaining 10 studies, several had very small samples (<10), meaning that the study could have included only one or two teachers of color.

Of these 104 studies, the vast majority of studies was conducted by researchers in their home universities. In fact, only eight studies explicitly studied PSTs from contexts outside of the researchers' own institutions. There is a small number of studies in which the researchers do not explicitly identify the research site as their home university, but there typically are hints that, in fact, the home institution was the object of study.

What Are the Methods That Have Been Used?

Nine of the 104 studies included 100 or more participants, including two that used large data sets with more than 1,000 teachers. The remaining studies were divided across three different sample size groups—less than 10 participants (31 of 104 studies), 10–30 participants (29 of 104 studies), or 30–100 participants (26 of 104 studies). Sample sizes for the remaining studies were not clear. More than three-fourths (83 of 104) of the studies used convenience sampling. Another 14 reported use of purposive sampling. Two large-scale studies used the Teacher Education and Development Study in Mathematics (TEDS-M) data that was acquired using a multi-stage sampling strategy (see Tatto et al., 2012). Another study randomly assigned PSTs to experimental conditions. The remaining four studies could not be classified.

With a small number of exceptions, studies of PST learning and development are largely descriptive and exploratory, typically focused on a single EPP. Approximately two-thirds of the research studies involved case studies and/or mixed methods. Often the line between mixed methods and other methodological approaches was blurry. For example, some studies (e.g., Sun & van Es, 2015; Yeh & Santagata, 2015) employed multiple methods within a quasi-experimental design that compared treatment conditions. Other studies (e.g., Swars, Smith,

Smith, Carothers, & Myers, 2018) could have been classified as case studies. The methods used in these studies consisted predominantly of interviews, observations, surveys, assessments, and document analysis. The vast majority of cases involved single programs within one EPP institution, though sometimes up to three EPPs were studied.

A second methodological approach that is used to study PST learning and development involves surveys. Twelve studies involved surveys. However, only three used a systematic survey sampling approach (Jacobson, 2017; Qian & Youngs, 2016; Ronfeldt et al., 2013). The rest relied on convenience sampling, most typically of PSTs in the researchers' home institutions. The modal number of institutions involved in these convenience sample studies was one. Survey instruments were almost always developed by the research team and were often used as pre- and post-test measures to study the impact of an intervention of some type. Only Qian and Youngs reported international survey results using TEDS-M.

The final methodological approach that was used with some frequency (seven studies) involved experimental or quasi-experimental approaches. Most of these were experimental and involved the random assignment of students and/or sections to a treatment or control condition. The vast majority of these studies was conducted at single institutions. A representative example of this type of study is Olson et al. (2016). Nine elementary methods courses at one institution were split into three treatment conditions to study the effects of unit planning and video analysis on PSTs' ability to understand and analyze samples of science teaching that they observed.

The vast majority of studies developed their own instruments in the form of surveys, interviews, observations, and assessments. In most studies using surveys, observations, and assessments, very minimal, if any, information is provided about the accuracy, reliability, or validity argument for the study measures. In addition, while occasionally studies made use of instruments that others had developed, the dominant approach for each study was to develop its own data collection protocols and analytic tools (e.g., Ghousseini & Herbst, 2016). For example, many studies (e.g., Adams & Gupta, 2017; Amador & Carter, 2018) developed their own coding schemes of some form of observational data, typically using open coding (Corbin & Strauss, 2008) and grounded theory (Glaser & Strauss, 1967) approaches. While many of these studies report using analytic methods appropriate for qualitative data analysis, these interpretive methods do not provide for estimates of accuracy or reliability that can be evaluated across studies. One recurring measure that has been used and that has a substantial literature about its measurement qualities is the self-efficacy survey developed by Tschannen-Moran et al. (1998).

Taken together, the most common research approaches to understanding STEM PST learning and development can be characterized as exploratory, relatively idiosyncratic, and local. There were, however, seven experimental or quasi-experimental studies. When compared to previous reviews of the TE research (Cochran-Smith & Zeichner, 2013), this is a notable shift. The vast majority of studies involves systematic interrogation of one or a very limited number of sites by researchers who are part of the EPP being studied. Many of these studies provide descriptions, sometimes very rich descriptions, of PST learning about certain constructs associated with teacher preparation and teacher learning. This work is most typically done within single and/or unique contexts.

The fact that measures are developed for each study often means that there is limited information on their reliability, robustness, or validity. Because there are few instances of common instrument use, coupled with the lack of any representative sampling within or across studies, the ability to generalize findings across studies is particularly challenging.

What Has Not Been Studied, and What Methodological Approaches Would be Needed?

The majority of studies reviewed for this paper fall under this section's purpose. There are multiple studies that address major questions of PST learning and development with respect to knowledge, practices, and beliefs. Yet, there are significant gaps that preclude a coherent understanding of how PSTs learn to become teachers of STEM.

First, within the journals that were reviewed, there are no studies of PST learning and development in engineering or technology.

Second, there are almost no studies that examine PST learning across different contexts. Studies do not examine learning issues across subject areas, grade levels (elementary vs. secondary), or even across different contexts such as institutions. There are almost no studies that examine how particular preparation practices are related to PST learning within or across institutions.

Third, there is little evidence that PST learning is examined across studies with common measures or even locally developed measures that have been validated for their use in particular studies. The implication of this is that there are no groups of studies that build on each other by examining similar constructs over different research contexts with common measures.

Fourth, there are very few research designs that examine PST learning in ways that lead to any generalizations beyond the particular case that is the subject of the study. The vast majority of survey and observation studies makes use of convenience or purposive samples, making it inappropriate to generalize findings beyond the sample studied. The relatively few experimental studies were done at single institutions with very small sample sizes, also making it inappropriate to generalize any findings beyond the particular context.

Taken together, the field is lacking a set of coordinated studies of PST learning that can build on each other, what Zeichner calls programs of research (Zeichner, 2013). There are very few complementary sets of studies that share the problems they address or the measures they use. While there is a reasonably well-developed literature of specific cases of PST learning, there have not been studies that lend themselves to develop broader and shared understandings of PST learning in STEM teacher preparation.

In order to move the field forward, a set of complementary research efforts that builds on and leverages existing work is necessary. The fundamental questions being asked about PST learning and development are reasonable (if broad) and derive from the research and policy literature. The fundamental challenge will be to develop studies that can explore critical issues with designs, samples, and instrumentation that allow for generalization to larger and more representative groups of PSTs and programs. Another fundamental challenge will be to have teacher educators and their close colleagues select and build programs of research that build one another's insights and can accrue across the relatively fractured body of research that now exists.

There are a number of ways that our understanding of PST learning can benefit from more sustained and concentrated efforts that are intentional about how they relate to each other. For example, the use of common conceptual frameworks that are aligned with current guidance documents (e.g., the Common Core, Interstate New Teacher Assessment and Support Consortium [INTASC] teaching standards) about core ideas such as reformed instruction, inquiry, and teacher learning can facilitate an understanding of findings across studies. If common measures are developed to focus on the aspects of teaching that cut across STEM subjects, the field can make progress in understanding the consistencies and differences in PST learning across different content or between elementary and secondary PSTs. Such common measures would compel the field to articulate more common definitions, which, if done over time through scholarly investigation and debate, could move the field forward. Work that cuts across institutions and makes use of systematic sampling strategies can lead to generalizations that go beyond most current work. Similarly, there may be interventions of common interest across institutions that can provide the basis for experiments that can support generalized causal inferences about PST learning.

Improving Educator Preparation Programs (EPPs)

This section refers to research that is focused on the improvement of STEM EPPs as institutional entities. This is distinct from the improvement of particular course-level pedagogical or curricular improvements that were discussed in the prior section.

What Are the Important Questions and Phenomena Being Addressed?

Strikingly, our search yielded virtually no published research that explored issues related to the improvement of EPPs. The only study classified as addressing this issue was a conceptual piece by Windschitl and Stroupe (2017) that argued for the redesign of teacher education practices in ways that would support preparation for teaching to the Next Generation Science Standards (NGSS).

Despite redesign efforts underway (e.g., Woodrow Wilson Academy of Teaching and Learning, Trellis) and broad policy calls for redesigning teacher education (e.g., Bybee, 2014; Davis & Boerst, 2014), there does not appear to be any systematic research efforts within the literature we reviewed examining the redesign of EPPs and its impact on teaching and learning in STEM fields.

The Sample: Who is Being Studied, and Who is Doing the Research?

The one article in this area was written by individuals at a research-intensive institution but is intended to provide direction for teacher preparation institutions broadly.

What Are the Methods That Have Been Used?

The scholars writing about this purpose use logical analysis and build on existing literature to make the argument for how program practices might address the NGSS.

What Has Not Been Studied, and What Methodological Approaches Would be Needed?

STEM teacher preparation does not stand alone in lacking an empirical base of research addressing EPP redesign. While a number of models of redesign (Ball & Forzani, 2009; Darling-Hammond & Bransford, 2007, Grossman, Hammerness, & McDonald, 2009) have been put forth more generally in teacher education, these documents have generally articulated a vision for—not data on—system redesign. There is very little published research that systematically explores the impact of such changes at the institutional level.

Conducting research to address EPP improvement necessarily implies the study of interventions broadly conceived. Programs have to decide to change what they are doing (i.e., they must intervene in the status quo). Such intervention happens frequently. Case studies and other methods can be used to describe and understand what happens during the course of these interventions while experimental and quasi-experimental approaches can be used to study the impact of such interventions on teaching and learning both in the EPPs and in the K–12 schools that eventually hire these new teachers. While new research designs are needed, the far bigger challenge concerns structural practices that make it very difficult to both establish and make changes to what happens inside EPPs.

As Windschitl and Stroupe (2017) note, any attempts to improve practice require identifying and then modifying practices. Yet, it is problematic to establish how EPP courses and programs are actually designed. As noted in an NRC report (2010) that summarized existing research on teacher preparation, "[t]here is little firm empirical evidence to support conclusions about the effectiveness of specific approaches to teacher preparation" (p. 4). While there have been studies about the relative effectiveness of EPP pathways (e.g., traditional vs. alternative-route) (e.g., Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009), characterizing what goes on inside of programs has long been challenging (NRC, 2010). The makeup of courses and programs is highly variable from institution to institution, making it difficult to conceptualize a sound cross-institutional, cross-program research design.

Nevertheless, the NRC (2010) report does identify a consistent set of dimensions of teacher education that is characterized as consistently important across all policy and research reports that have attempted to examine teacher preparation:

1. program purpose;
2. requirements for subject-matter knowledge;
3. requirements for pedagogical and other professional knowledge;
4. field and clinical experiences; and
5. faculty and staff qualifications. (p. 44)

Each dimension offers opportunities for systematic examination of core aspects of EPPs. A range of research designs, including experimental and quasi-experimental studies, could be carried out provided there was clarity about and control of the dimensions being studied. Clarity about these dimensions must also take into account relationships with and implications for the K–12 schools in which PSTs are placed. For example, a set of institutions could agree to some delimited set of requirements about STEM teacher preparation and then participate in some form of a mutually agreeable intervention. This would allow for systematic study over time. Alternatively, programs might use their own accreditation data as the basis for studying program improvement with respect to these dimensions. Absent clarity about and control of what programs are actually doing, it is unlikely that any firm claims can be made about factors that contribute to particular STEM EPP outcomes.

Contributing to EPP Accountability

This section refers to research focused on accountability efforts in teacher education. There have been numerous calls for increased accountability of teacher education programs (CAEP, 2016a; Levine, 2006). The NRC (2010) report on teacher preparation noted that despite the calls for accountability, the evidence about the utility of accountability efforts and particular types of measures was very limited. Consequently, the report called for significant research to address these issues.

What Are the Important Questions Addressed?

Despite the NRC (2010) report, there has been almost no research done on accountability of STEM teacher preparation programs. Our search yielded two published documents, neither of which met our criteria of being STEM-focused. One (AERA, 2015) is a statement put out by the American Educational Research Association urging caution in using value-added models to evaluate EPPs.

The second article (Ronfeldt & Campbell, 2016) examined the utility of using teacher evaluation observation scores of practicing K–12 teachers to differentiate the quality of EPPs but was not specifically focused on STEM.

The Sample: Who is Being Studied, and Who is Doing the Research?

Ronfeldt and Campbell (2016) was based on a secondary analysis of statewide administrative data that included 9,500 EPP graduates.

What Are the Methods That Have Been Used?

The Ronfeldt and Campbell (2016) study employed a quasi-experimental method that attempted to identify the causal influence of particular EPPs. Using a variety of statistical modeling approaches, Ronfeldt and Campbell argued that, indeed, observational scores could be used to distinguish among preparation institutions and that these distinctions were positively related to value-added estimates. Teachers from all grade levels and subject areas were included. Programs were defined by their institution, whether the program was graduate or undergraduate, and whether the teaching endorsement was for elementary, secondary, or special education. Data were not disaggregated by content field. It is of note that, despite not appearing in the journals included in this review, there are studies of EPP accountability that appear in the literature on the economics of education (e.g., Goldhaber, Liddle, & Theobald, 2013; Koedel, Parsons, Podgursky, & Ehlert, 2015).

What Has Not Been Studied, and What Methodological Approaches Would be Needed?

Echoing the NRC (2010) recommendations, there needs to be research focused on the relationship of STEM teacher preparation and different types of K–12 student outcomes, including those that go beyond standardized achievement measures (Cappella, Aber, & Kim, 2016), in order to establish reasonable and productive accountability processes. Such processes should take into account the cautions raised by Rowan and Raudenbush (2016) concerning the ways in which accountability pressures can actually lead to significant unintended consequences in institutional behavior. For example, cooperating teachers (CTs) might choose not to have student teachers because they are concerned that their own students' achievement and, consequently, their own growth scores might drop in the district's accountability system. Thus, there also needs to be research exploring the effect of imposing accountability measures and practices and its impact on EPPs and the K–12 institutions they serve.

In some professions such as nursing, the program accreditation process serves as an accountability pressure. We did not find any articles that investigated how program accreditation operates to shape STEM teacher learning. This is an area ripe for research.

Describing and Understanding Relationships Between STEM Teacher Preparation Programs and Other Valued Outcomes

What Are the Important Questions and Phenomena Being Addressed?

The 22 studies in this category were roughly evenly distributed across a range of foci that operates at higher levels of aggregation all the way down to specific narrow components of programs. Specifically, the studies can be grouped as follows: country-level policies that shape teacher preparation (one study); whether the program was an “alternative” or “traditional” certification program (five studies); institutional features of teacher preparation programs (five studies); program components (six studies); and student teaching (five studies). Only 12 of these studies met the criteria for being STEM-focused.

At the highest level of aggregation, one international study considered the relationship between the degree to which a country provided quality assurance of teacher preparation programs and graduates (e.g., the strength of the accreditation policy for EPPs, requiring an undergraduate degree in mathematics prior to certification) and graduates’ levels of CKT and mathematical pedagogical content knowledge (Ingvarson & Rowley, 2017).

At the program pathway level, five studies considered the relationships between alternative, traditional, in-state program completers, and out-of-state program completers and valued outcomes such as practicing teacher effectiveness (Bastian & Henry, 2015; Shuls & Trivitt, 2015); teacher retention and mobility (Hansen, Backes, & Brady, 2016; Redding & Smith, 2016); and the diversity of the teacher workforce (Marinell & Johnson, 2014).

Moving down a level of aggregation, five studies focused on institutional program level characteristics or features of programs (e.g., types of coursework, content area partnerships with schools, size of program, and requirements for senior seminars or early field placements) and teacher effectiveness (Lincove, Osborne, Mills, & Bellows, 2015; Preston, 2017); pre-service teacher engagement in their programs (Kim & Corcoran, 2017); the development of beliefs, knowledge, and practices over time (Swars et al., 2018); and the development of undergraduates' interest in becoming a teacher (Swanson & Coddington, 2016).

Another six studies sought to describe and understand how the knowledge and beliefs (e.g., perceptions of courses, teacher educators' views of ethics, and the nature of science) developed within and across program components were related to teachers' classroom practices (Hiebert, Miller, & Berk, 2017; Morris & Hiebert, 2017), teachers' beliefs (Bahr, Monroe, & Eggett, 2014; Herman & Clough, 2016), opportunities for PSTs to learn about ethics (Maxwell et al., 2016), and K–12 student learning (Shaw, Lyon, Stoddart, Mosqueda, & Menon, 2014).

At the lowest level of aggregation, five studies explicitly focused on the relationships between student teaching and other valued outcomes related to the first years of teaching (e.g., teaching effectiveness, K–12 student achievement, self-efficacy, or the location of the school in which the candidate accepted employment). Only one study looked at the relationship between student teaching and prospective teacher knowledge, beliefs, or capabilities. That study considered the relationship between the length of student teaching and beginning teacher self-efficacy (Clark, Byrnes, & Sudweeks, 2015). Two studies considered the relationship between structural features of both teacher preparation and its outcomes (Goldhaber, Krieg, & Theobold, 2017; Ronfeldt, 2015), investigating the relationships between the student teaching placement characteristics (i.e., teacher collaboration, student achievement gains, teacher retention, match to first-year placement) and first-year teaching effectiveness as defined by VAM. One study considered CTs' perceptions of the impact of student teachers on their K–12 students' achievement (Tygret, 2017). The last study considered the relationships between the geographical location of student teaching placements and the geographical location of pre-service teachers' EPPs, homes, and first-year job placements (Krieg, Theobald, & Goldhaber, 2016).

Who is Being Studied, and Who is Doing the Research?

The 22 studies in this group covered all grade levels from K–12; half addressed both elementary and secondary grades, eight focused on elementary, and just three focused exclusively on secondary grades. STEM subjects as well as non-STEM subjects were represented in these studies. Mathematics was the most frequently included STEM subject (19 studies) with science included in nine studies. Due to the number of large-scale studies that used state or other large-scale institutional data sets, almost one third of studies covered all subjects. The teachers' racial backgrounds are unclear because there is a good deal of missing information. Ten of the studies did not report the racial background of the teacher participants. Of the 12 studies that did report teachers' racial backgrounds, 11 had predominantly White participants (i.e., more than 80% of the study's sample reported they were White). There was, however, one study that collected data from a larger proportion of teachers of color, with roughly 80% identifying as African American or Latino.

Researchers in ten of the studies did not create or oversee the creation of the study's data as a result of their organizational roles. In general, these studies used large-scale state or national data. In nine studies, the researchers seemed to produce the study's data as part of their professional organizational responsibilities (e.g., being the principal investigator [PI] of a grant that was the intervention, leading the university-school district partnership, teaching a course in which data were collected). It was unclear what role researchers played vis a vis the data and their home organizations in the final three studies.

What Are the Methods That Have Been Used?

More than half of the 22 studies included 100 or more participants, with eight of those studies using large data sets with more than 1,000 teachers. The remaining studies were divided across three different sample size groups—less than 10 participants (three studies), 10–30 participants (four studies), or 30–100 participants (three studies). Samples for 11 studies were designed to reflect a population of teachers through a deliberate sampling strategy (e.g., all teachers at five Washington EPPs, all North Carolina teachers, all beginning teachers who had adequate data in a district). Another ten studies used convenience samples. One study used a purposive sample—deliberately seeking out variation in participants based on the study's research questions. Because more than half of the studies were large-scale studies, those studies allowed for generalizations at the district, state, or national level. Nine studies were specific to a single EPP.

Two methodological approaches dominated this literature: survey methodologies (eight studies) and quasi-experimental regression-based analyses (seven studies). In addition, there were a small number of institutional or program case studies (two studies), large-scale descriptive database analyses (two studies, using the *Schools and Staffing Survey* [SASS; see https://nces.ed.gov/surveys/sass/]), and mixed methods (three studies).

Due to the number of studies that used existing databases, many of the studies relied on measures created by others. For the studies that collected primary data, the researchers created their own measures, and little was reported about basic aspects of those measures, which were often surveys and, to a lesser degree, original assessments of various types (e.g., Hiebert et al., 2017; Morris & Hiebert, 2017).

Understandably, researchers tailored their survey and assessment measures to their specific study's purpose. Accuracy of measures was not generally a significant issue in instrumentation because the vast majority was surveys or other instruments that were subsequently qualitatively coded. However, there was generally little information about the reliability of the instruments or scales within instruments, and because the study was reporting on the first use of the measure, the instrument did not have a developed validity argument. Finally, the instruments were clearly aligned to the study's purposes or the program, course, or intervention under investigation, but it is unclear how those specific programs, courses, or interventions are related either to the EPP's goals or the larger community's goals (e.g., the Common Core State Standards or the NGSS). It seems likely they are related, but this is not specifically delimited in the articles. One article, however, was an exception: Clark et al. (2015) modified an existing survey that had been used once by other researchers, the *Total Quality Partnerships Teacher Survey* (Lasley, Siedentop, & Yinger, 2006), though the revision was significant enough to give it a new name.

What Has Not Been Studied, and What Methodological Approaches Would be Needed?

There are three significant unexplored or underexplored areas within this body of research—the first is a marked absence of program descriptions that are then linked to valued outcomes; a second is the relationships between parts of programs vis a vis one another and valued outcomes; and the final area is the lack of common and robust instrumentation around valued outcomes.

Research suggests that programs are complex—they are comprised of a variety of courses, field placements, and other experiences that frequently vary based on the sequence in which PSTs experience them as well as the particular syllabi and faculty or staff member leading PST learning (Grossman et al., 2009). The mathematics program at an EPP may be quite different than the science program at the same EPP. What is strikingly missing from the delimited body of research reviewed here are program descriptions. There are studies that carefully describe aspects of programs—a subject-matter partnership, the ways in which subject matter is taught to aspiring mathematics teachers—but these are not descriptions of the larger programs themselves. In order for the field to develop a robust understanding of how programs shape valued outcomes, we must share common understandings of what programs are and link those to outcomes.

On a related point, the current body of research reviewed here does not help us understand how particular parts of programs contribute to valued outcomes given the rest of the program. We do not know, for example, whether the sequences of courses or experiences within a program matter to the development of PSTs' CKT or orientation to reform-minded instruction. If a faculty member developed a new CKT course for elementary school teachers and wanted to insert it into a program's course sequence, the field does not have designs or detailed descriptions of programs that could guide that faculty member's decision or a study of the course in the context of PSTs' other courses, prior knowledge, or experiences. In other words, studies show that there are relationships between sub-components of programs and outcomes, but we have little information on other sub-components so we cannot contextualize those findings to determine their meaning at a program level. To summarize, the designs and approaches used in this literature do not cohere to develop the field's understanding of how programs function and how they relate to a broad range of valued outcomes.

Finally, the lack of common instrumentation also contributes to the challenge of developing a strong field-level understanding of how programs shape valued outcomes. Many programs are accredited and, depending on the details of that process, it is possible that the field might use accreditation requirements to develop a much more robust set of instruments to document valued outcomes. For example, some EPPs are using common observation tools to evaluate PSTs in student teaching placements. Such tools could be developed and validity arguments created for their accuracy, reliability, general psychometric quality, and relationships to various outcomes that the EPPs and/or the program values. The same might be true of various surveys of valued constructs. For example, cross-program and EPP use of instruments that measure the nature of science, STEM self-efficacy beliefs, or CKT would be useful. Such common instrumentation would allow researchers to compare within and across programs to better describe and understand programs and their components as related to these valued outcomes.

Understanding Assessments and Measurement of STEM Teacher Preparation Quality

What Are the Important Questions and Phenomena Being Addressed?

There are many measures used in preparation programs to document pre-service teachers' knowledge (e.g., Akerson et al., 2017), practices (e.g., Hiebert et al., 2017), and beliefs (e.g., Trauth-Nare, 2015) before, during, and after preparation. These measures can vary widely from survey instruments to multiple-choice assessments of knowledge to observation ratings during student teaching. Despite the abundance of measures used in studies and this wide variation in type of measures, there were only five studies whose purpose was to better understand the validity and reliability of measures of teacher preparation quality; and those studies did not cover the range of measures used to measure important constructs in teacher education—in particular, research on beliefs and knowledge were not present or very limited. Further, questions were aimed at a high level of aggregation and not framed or reported narrowly around STEM.

Four of the five focused on measures of teaching practices (Bastian, Henry, Pan, & Lys, 2016; Bryant, Maarouf, Burcham, & Greer, 2016; Caughlan & Heng, 2014; Ronfeldt & Campbell, 2016), and one focused on knowledge measures as predictive of later performance in teacher education (Evans, 2017). Of the four performance measure studies, three (Bryant et al.; Caughlan & Heng; Ronfeldt & Campbell) focused on the construct validity, predictive validity, or utility of observation scores for different purposes. The fourth study, Bastian et al., considered the quality of Teacher Performance Assessment (TPA) portfolio assessment scores created by local institutions—investigating construct validity, reliability, accuracy, and predictive validity. The final study in this group investigated the relationships between scores on screening measures of PST knowledge (grade point average [GPA] and Graduate Record Examinations [GRE®] scores) prior to entry into teaching and performance in EPPs as measured by preparation program GPAs.

Who is Being Studied, and Who is Doing the Research?

All but one study drew on data from all subjects (STEM and non-STEM) and all grade levels. Caughlan and Heng's (2014) study critically reviewed the language in three observation protocols that are used across grades and subjects and, therefore, did not draw on participant data.

Apart from Caughlan and Heng (2014), only two of the remaining four studies reported on participants' racial or ethnic backgrounds. Of the two that did report racial background information, one study was carried out with predominately (88%) White PSTs, and one had a somewhat larger proportion (approximately 20% of the sample) of African American teachers.

All of the researchers had some connection to the organization that carried out the data collection or could have used the study's results to inform organizational decision making (although the degree to which that decision making was actually informed by study results is unclear). In three studies, at least one author worked for the organization that produced the data, and the analyses had connections to their job responsibilities (i.e., they worked in the program). In a fourth study, the authors (Ronfeldt & Campbell, 2016) collaborated with a state that was interested in determining how best to evaluate EPPs, and the authors were responsible for informing the state about the quality of observation scores for that purpose. In the study with the most distant connection to the data collection and results, the authors (Bastian et al., 2016) used data collected from another collaborating EPP, although the EPP was interested in revising the way it carried out PST assessment, thereby making the study's results relevant to the work of the collaborating EPP.

What Are the Methods That Have Been Used?

This group of studies generally used large-scale assessment methodologies. Of the studies that relied on participant data, all included 100 or more participants, and one made use of data from more than 9,000 teachers. Three studies' samples came from single EPPs and were convenience samples. The fourth study was an entire state's EPP system (44 institutions and 183 programs in Tennessee) and, therefore, involved the total population sampling of one state's programs.

As might be expected, this research focused on construct and predictive validity concerns, using traditional analytic approaches such as factor analyses and construct mapping as well as correlational and regression analyses to determine the relationships between measures and predictive outcomes such as performance in the preparation programs or in-service measures such as observation scores, VAM estimates, and labor market outcomes.

What Has Not Been Studied, and What Methodological Approaches Would be Needed?

Recent observation research on practicing teachers suggests that measures of teaching quality are sensitive to students' prior academic achievement of K–12 students (e.g.,Whitehurst, Chingos, & Lindquist, 2014), subject matter taught (Bill and Melinda Gates Foundation, 2012), grade levels (e.g., Mihaly & McCaffrey, 2014), and the observation systems used to create scores (Liu, Bell, Jones, & McCaffrey, 2019). Studies in this group were not STEM-focused. They did not focus exclusively on STEM teachers or subjects; nor did they focus on specific grade levels or generally take account of the impact various groups of K–12 students might have had on the portfolio or observation scores. We would expect that if measures to evaluate K–12 teaching are sensitive to the students, grades, and subjects under consideration, using the same measures to evaluate pre-service teaching will also be sensitive to these factors. At a minimum this should be an object of study.

The studies in this group did not sufficiently disaggregate so that readers could learn much about the quality of the knowledge or practice measures for STEM. Additional disaggregation and targeted studies that illuminate how various measures function for STEM subjects, grades, and populations of K–12 students would be useful. For example, what is the relationship of content knowledge and PCK to measures of practice and student outcomes across STEM disciplines and as compared with other fields?

One very clear area for additional scholarship concerns assessment practices that have been understudied. First, although research suggests that PSTs' beliefs, such as self-efficacy, beliefs about evolution, reform-oriented mathematics, and the nature of science, are related to important teaching practices in complex ways (Mansour, 2009), there were no studies about belief instruments used in STEM preparation.

Second, there was minimal research on the use of high-stakes portfolio assessments, a form of assessment that plays an increasingly large role in the preparation of STEM teachers. Much has been raised about both the strengths (e.g., Darling-Hammond & Hyler, 2013; Lahey, 2017) and weaknesses (e.g., Tuck & Gorlewski, 2016) of high-stakes portfolio assessments; however, we found only one study. In our unstructured search, there was an additional study using Washington state teacher data that found a strong predictive relationship between edTPA scores and teachers' probability of getting hired. There was some evidence of a predictive relationship between edTPA scores and value-added measures of teacher effectiveness, depending on the modeling assumptions and subject matter (Goldhaber, Cowan, & Theobald, 2017).

Finally, there was little research that examined the range of measures used to accredit programs. Accreditation often uses a wide range of measures (e.g., portfolios, observations, knowledge measures) to document PST learning. Recent changes to national accreditation standards (to which measures must be aligned) have resulted in a great deal of discussion (Cochran-Smith, Stern et al., 2016). Depending on the accrediting body, STEM programs may be working toward different teaching standards than non-STEM programs. Research can assist in better understanding how to best focus accreditation measures in STEM fields such that they support the development of strong beginning STEM teachers.

Framing or Reframing Issues of STEM Teacher Preparation

What Are the Important Questions and Phenomena Being Addressed?

Together the 29 articles in this group carried out original scholarship whose purpose was to frame or reframe specific issues in teacher preparation. Only eight of those articles were STEM-focused. Authors accomplished their reframing using different approaches. Nine of the articles reviewed empirical and/or scholarly literature on a specific topic in order to provide a new framework that might substantively guide future research (e.g., a 4-task framework that would support emphasizing equity in teacher education [Cochran-Smith, Ell et al., 2016]). Some of these literature reviews provided substantive insights about a specific aspect of teacher preparation (e.g., the state of the field, the factors that shape teacher resilience, major programs of research, and historical changes in the field over time). For example, two studies (Sleeter, 2014; Özçınar, 2015) reviewed large numbers of articles to determine the degree to which the field is coherent or is carrying out empirical studies likely to lead to the insights necessary for improvement. Özçınar did this by carrying out a co-citation analyses, and Sleeter carried out a more traditional literature review with careful attention to the proportion of methodologies used by researchers that can lead to robust research insights.

The second most common methodological approach used in 7 of 29 of the studies was to investigate specific teacher preparation practices and then either draw implications for preparation or offer new ways to think about preparation or research on teacher preparation. Practices varied widely from the use of simulations (Dotger, 2015) to the use of co-teaching (Baeten & Simons, 2014) to programmatic approaches to practice-based teacher preparation (Janssen, Grossman, & Westbroek, 2015).

A smaller group of studies (5 of 29) carried out a literature review, brought new literatures together, or carried out logical analyses to argue for the utility of specific pre-service pedagogies or epistemological perspectives on preparation (Avraamidou, 2014; Kahn & Zeidler, 2016; Shane, Binns, Meadows, Hermann, & Benus, 2016; Steele, 2016; Windschitl & Stroupe, 2017). Windschitl and Stroupe carried out a logical anlayses to argue that teacher educators should use powerful principles derived from the research underlying *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (NRC, 2012) to inform the design of courses and preparation experiences rather than more traditional alignment models of implementing new standards (e.g., taking the NGSS, aligning them to K–12 student learning expections, then designing instructional activities).

Four articles (Bullough, 2014; Fuller, 2014; Paufler & Amrein-Beardsley, 2016; Zeichner, Payne, & Brayko, 2015) critiqued taken-for-granted assumptions, policies, and practices that were generally connected to accountability issues in teacher preparation but not necessarily focused on STEM. For example, Fuller's analysis illuminated concerns about an influential and contentious annually published policy report—the National Council on Teacher Quality's (NCTQ, 2013) report about teacher preparation quality—documenting flaws within the report's methodlogical approach. Zeichner and colleagues critiqued the underlying assumptions on which preparation programs' relationships with K–12 schools are based—assumptions about who holds knoweldge and expertise relevant to training teachers. The other two studies critiqued, first, the NRC's underlying assumptions about the nature of teaching as well as who and what matters with respect to improving teaching and, second, how exisiting conceptualizations of accountability should be revised to empower teachers as change agents, value education, and support preparation partnerships. None of these concerned STEM exclusively.

Historical analyses that trace specific aspects of teacher education, research on teacher education, and/or teacher policy over time make up another small group of three studies

(Cochran-Smith, 2016; Darling-Hammond, 2016; Forzani, 2014) that are not STEM-focused specifically. Both Cochran-Smith and Darling-Hammond use AERA presidential addresses to answer questions about how the discussion of and research on teaching and teacher education evolved.

Finally, there was one international study of Teach for India and Teach for America (Blumenreich & Gupta, 2015) that used institutional discourse analysis to illuminate and question the appropriateness of the programs given their cultural contexts. Again, this was not exclusively STEM-focused.

Who is Being Studied, and Who is Doing the Research?

Articles in this group were focused on field-wide conceptual issues and, therefore, generally did not specify the grades and subjects to which they are meant to generalize. However, for the most part, they are meant to apply to all grade levels, and authors reviewed literatures from a wide range of grade levels. Articles also generally apply to all STEM (and other) subjects; however, there was one study specific to mathematics and seven that were specific to science. Finally, the arguments presented were synthetic in nature and, with one exception (Felton-Koestler, 2017), did not collect primary data from participants.

What Are the Methods That Have Been Used?

As previously described, there was generally no primary data collection in this group of studies. The methodological approaches included conceptual syntheses and frameworks (23 of 29); historical analyses (3 of 29); policy case studies (1 of 29); and discourse analyses (2 of 29).

What Has Not Been Studied, and What Methodological Approaches Would be Needed?

The methodological approaches that have been used among these studies are appropriate to the claims and inferences drawn from them. This is not to say that other methods would not be appropriate. For example, one might carry out primary data collection from a design experiment about how to improve teacher preparation programs and argue for a new way to think about the improvement of preparation programs. Such methodologies might provide additional and/or stronger warrants for future teacher preparation directions. Even if the current methodologies remain the dominant form of inquiry, as teacher education research addresses a broader set of issues, many identified throughout this report, new approaches to reframing teacher preparation are likely to emerge. In the discussion section to follow, we consider some of these reframing issues.

Several issues stand out as needing further consideration. First, the field would benefit from considerations about framing and reframing teacher preparation with specific reference to STEM. The context of STEM preparation does vary in some important ways from other fields. For example, potential teachers in STEM fields, more than those in many other academic disciplines, are in high demand. Such individuals are more likely to have competing employment opportunities than potential teachers in the liberal arts. Additionally, there are more shortages of teachers in STEM fields than in many other academic fields. Many STEM fields also are less likely to include women and people of color as PSTs. While these issues are being discussed within teaching more generally, research is needed to better conceptualize how to increase the racial and language diversity of those entering the field of STEM teaching.

Work that interrogates how novice STEM teachers learn the knowledge, practices, and beliefs they use in their classrooms is also needed. Finally, building on critical and socio-cognitive perspectives, it would be helpful to develop frameworks and literature reviews that allow us to better understand what we know and do not know about preparing STEM teachers for diverse learners. This suggestion is consistent with two recent large-scale reviews of research on mathematics and science teaching (Chazan, Herbst, & Clark, 2016; Windschitl & Barton, 2016).

Understanding STEM Teacher Educators and Their Practices

What Are the Important Questions and Phenomena Being Addressed?

Of the 14 studies considering STEM teacher educators and their practices, eight were STEM-focused. Across all of the studies, most (12 of 14) investigated university- or EPP-based teacher educators; however, there were two studies (Gareis & Grant, 2014; Hoffman et al., 2015) that specifically focused on cooperating teachers. One longitudinal study (Gareis & Grant) that drew on 10 years of data from a training program for CTs in Virginia investigated how the self-efficacy of trained CTs differed from that of untrained CTs. The researchers also looked at the shorter- and longer-term self-reported outcomes for novice teachers mentored by trained and untrained CTs (e.g., quality of student teaching experience, desire to remain in profession). Another study (Hoffman et al.) was a literature review of coaching practices used by CTs. There was a third study (Hjalmarson, 2017) that involved the training of K–12 school-based mathematics coaches; however, the object of study was the teacher educator’s decision making in designing the course for the coaches rather than the coaches themselves.

The studies that involved university- or EPP-based teacher educators can be grouped into a focus on the knowledge, practices, and beliefs of teacher educators. In contrast to the studies of PSTs’ learning and development, the studies of teacher educators largely focused almost exclusively on their practices; only one study (Castro Superfine & Li, 2014) focused on the knowledge demands necessary for teaching mathematics content to pre-service teachers. In that study, the authors reviewed course artifacts from 10 iterations of a course (taught by various instructors) to determine the nature of the knoweldge that teacher educators need to teach mathematics content.

The other 11 university- or EPP-based studies focused either on relatively narrowly defined, delimited practices or on broader, less well-defined practices. Narrower practices included using metaphors to help PSTs learn curriculum (Lynch & Fisher-Ari, 2017), using rehearsals (Davis et al., 2017; Kazemi, Ghousseini, Cunard, & Turrou, 2016), using video (Christ, Arya, & Chiu, 2017), and using a post-observation conferencing tool for field instruction (Soslau, 2015). Broader practices included teacher noticing (Amador, 2016); collaboration processes between mathmaticians and mathematics educators (Bleiler, 2015); engagement in the policy process (Tuck & Gorlewski, 2016); or the design of courses (Hjalmarson, 2017; Li & Castro Superfine, 2018). There was one article (Goodwin et al., 2014) that described teacher educators' perceptions of the nature of their roles and their preparedness to carry out those roles.

Who is Being Studied, and Who is Doing the Research?

Similar to other purposes, studies in this group covered all grade levels and subject matter. Specifically, the teacher educators under consideration worked with elementary grades (four studies), secondary grades (two studies), both secondary and elementary grades (six studies), or did not report a grade-level specialization (two studies). As with several other purposes, mathematics was included almost twice as much as science: 79% versus 43%, respectively. There were two studies that did not report the subject-matter focus of the teacher educators being studied. Although the studies included teacher educators who were traditional university- or EPP-based course instructors, there were also studies that investigated cooperating teachers' practices. Unfortunately, none of the studies reported on participants' racial backgrounds.

Half of the researchers in this group created or oversaw the creation of their respective study's data. All of these seven studies were self-studies of some variety, although some included additional researchers (and co-authors) at various stages of the data collection (e.g., Davis et al., 2017). Three of the 14 articles' researchers were scholars using data to which they did not have any specific employment connection (e.g., surveying a wide range of teacher educators through a professional organization's listserv, carrying out a literature review). In four studies, there was not enough information in a given article to determine the role of the researcher.

What Are the Methods That Have Been Used?

The majority of studies in this group relied on small samples; 8 of 14 had samples of less than 10 participants. However, there were three large-scale studies with more than 100 participants. Two of the studies did not report the number of participants that contributed to the data, and one study was a literature review and, therefore, did not include direct data collection from participants. Consistent with these small sample sizes, nine articles used convenience samples, and just one used a purposeful sample. The remaining four did not need to sample either because it was either a literature review or a self-study of some variety.

Given the need for additional research on teacher educators of all types, much of the work in this area made use of methodologies that were foundationally descriptive or focused on understanding how specific practices or processes operated. The methodologies were diverse and included critical perspectives, mixed methods that relied on surveys and interviews, participatory action approaches, ethnographic approaches, and case studies. Six studies were cases of various types (e.g., teacher educator practices in courses, collaboration between a mathematics educator and a mathematician). The notable exception to this pattern was a literature review (Hoffman et al., 2015).

There were noticeably few details on the instruments used in many of the studies. Perhaps this is not surprising given the proportion of self-studies in the group and the reality that many self-studies use measures that do not have established validity, accuracy, and reliability data from other studies. Of the studies that we could determine used survey instruments, only one used a survey for which there were multiple uses of that survey instrument over time, as documented in other cited reports (Gareis & Grant, 2014). Further, the study that relied extensively on a survey (Goodwin et al., 2014) did not include a description of the survey's development, the instrument itself, or the instrument's reliability.

What Has Not Been Studied, and What Methodological Approaches Would be Needed?

There is a paucity of research on STEM teacher educators. In general, and particularly for STEM, there are three broad areas that would benefit from additional research: a deeper focus on the knowledge and beliefs of teacher educators of all types; a better understanding of the knowledge, practices, and beliefs of CTs; and more systematic building of the knowledge base regarding all teacher educators with increasingly robust and transparent methodologies.

This group of studies had no emphasis on teacher educators' beliefs and almost no emphasis on knowledge, narrowly defined—there was just a single study about the knowledge needed to teach PSTs mathematical content. Given the prominence of this work in considering teaching and teachers within the K–12 context, it is striking that research on teaching within the teacher preparation context is absent.

There was also very little emphasis on CTs, who play a critical role in the education of novice STEM teachers. There were only two studies that focused on CTs. As a group, CTs are important because they frequently see and participate in STEM teacher development on a more regular basis during PSTs' field placements than do university- and EPP-based teacher educators. Just like K–12 students' careers with multiple teachers, PSTs experience many different teacher educators, including CTs and other school-based teacher educators. The pattern or quality of those teacher educators may help us better understand STEM PST development; however, in order to develop that understanding, we must better understand who the teacher educators are and how they interact with PSTs. Additional studies can build on the considerable literatures and instrumentation already developed to study STEM mentor teaching and induction as well as the knowledge, practices, and beliefs of in-service STEM teachers. Future research might also consider how the "match" between cooperating teacher and PST might shape PST learning and development.

Finally, there were many studies that did not report important aspects of study design, most noticeably the teacher educators' racial backgrounds and the use of measures. Recent research in all subjects provides some limited support for the idea that teachers' racial backgrounds matter to student learning (Bates & Glick, 2013; Egalite, Kisida, & Winters, 2015; Gershenson, Holt, & Papageorge, 2016). It is possible this may matter for PSTs and/or for PSTs working in field placements with specific racial characteristics. If researchers do not track and report racial and/or ethnic background, the field cannot learn what role(s) they play in teacher educators' knowledge, practices, and beliefs. There was also very little focus in these studies on standardized measures. As studies in this area move from descriptive accounts and studies with small numbers of participants, it is important to develop measures that can be used in accurate, reliable ways that build validity arguments over multiple studies. This type of measure documentation and development is generally absent from studies oriented toward understanding teacher educators and their work.

Discussion

If one begins with core questions of interest that have been identified by policymakers, researchers, and practitioners, we find that there is much work to be done. While there are rich, well-developed bodies of research about STEM teaching based on in-service teaching (Chazan et al., 2016; Windschitl & Barton, 2016), the same cannot be said for teacher preparation based on our review and others' (Cochran-Smith & Zeichner, 2013; Özçınar, 2015) reviews. Overall, across purposes there were recurrent themes that characterize the current status of STEM teacher preparation research.

Focus

Research only addresses a small fraction of the kinds of questions that have been the subject of research and policy interests as captured in the seven research purposes identified. Almost 60% of the studies involved understanding STEM PST learning and development, yet there was very little research on goals that are the highest priority to policymakers, including accountability, measurement, and EPP improvement. Of the studies that did focus on these issues, few were STEM-focused. Additionally, there is relatively little sustained focus on the knowledge, practices, and beliefs of those who prepare STEM teachers, including teacher educators and cooperating teachers.

There is substantial analysis of particular pedagogical approaches and organizational structures within teacher preparation. For example, many studies of STEM PST learning and development describe how particular curricula or experiences shape PST knowledge, practices, and/or beliefs. Absent, however, are descriptions and studies of how those pedagogies fit into programs and combine to support PST learning outcomes. Also absent are understandings and critical analysis of STEM programs, which are prerequisite to developing research-based initiatives to improving those programs.

It is also striking that, within our search, we found no studies that addressed teacher preparation in two of the four STEM subjects—technology and engineering. There was substantially more work researching the preparation of elementary than secondary teachers—although the elementary work was generally not exclusively STEM-focused—and generally more research on mathematics than science, regardless of grade level.

Despite widespread concerns about and calls for access to high-quality STEM education for all students (see https://www.ed.gov/stem), there is very little research directly focused on STEM preparation for teachers of traditionally underserved K–12 students. Across all purposes, there was no focus on the specific nature of preparation when PSTs are going to teach traditionally disadvantaged K–12 students, though many studies took place with teachers who would eventually serve such students or were serving such students in their field placements. Indeed, the lack of attention to these issues is apparent in that a significant proportion of studies do not even report on the demographic characteristics of PSTs studied. For research to shed light on effective STEM teaching for all students, reporting the demographics of teachers and the students with whom they are working is only a very modest starting point. Much more attention must be given to how preparation influences understanding and practices of PSTs.

Finally, there was little research that can help us understand key issues across major distinctions in STEM teacher preparation such as subject matter, grade level, or the K–12 students with whom the PSTs will work. For example, how does teacher preparation differ across STEM fields? What distinguishes the preparation of elementary from secondary STEM teachers? How does preparation differ when EPPs are focused on preparing STEM teachers to work with students from underserved populations versus other populations?

Context

In summarizing the status of this body of research it is important to put STEM teacher preparation in its social and organizational context so as to more easily notice what research has not yet been carried out. An important, yet underdeveloped dimension of current research is the role of context; we do not know enough about what helps PSTs develop given specific contexts. If, for example, PSTs eventually teach in rural schools, are particular learning experiences more or less valuable for them? Are PSTs with certain experiences more in need of specific types of learning opportunities? The role that both the teaching and EPP contexts play in teacher education were generally not compared in these studies. Theorizing and empirical studies of how context—defined in many ways—matters would propel the field forward.

A second contextual issue is that teacher preparation is often thought of as synonymous with preparation programs. Conventional wisdom suggests that teachers learn what they need to know in their teacher preparation programs. But this is not true; and it speaks to how we define and study the teacher education context as a location for PST learning.

Teacher preparation includes many other institutions, perhaps most importantly, institutions of higher education (IHEs) and K–12 schools. These institutions are largely responsible for the subject-matter knowledge and the teaching practices that STEM teachers rely on once they are in their own classrooms. Through associate's and bachelor's degrees, IHEs teach PSTs almost all of the subject matter they need when they begin teaching. Much of the rest comes from the K–12 schools that teachers attended. For example, elementary STEM teachers learn about place value and evaporation in their own K–12 education. Secondary STEM teachers learn linear algebra or organic chemistry during their undergraduate subject-matter majors.[3] To take but one example, despite significant progress developing measures of teachers' CKT (e.g., Ball, Thames, & Phelps, 2008, Gitomer, Phelps, Weren, Howell, & Croft, 2014; Hill, Rowan, & Ball, 2005; Hill, 2010) we know little about how that knowledge develops over preparation institutions, nor whether it might be improved through different pedagogical approaches across higher education and K–12 institutions.

PSTs also learn specific teaching practices (e.g., carrying out a discussion, introducing a lesson) in teacher preparation programs. This learning often takes place in K–12 teaching contexts in which they interact with K–12 students and their cooperating teachers. While many of the specific learning opportunities afforded to PSTs are largely dictated by the K–12 students, curricula, and teacher colleagues in those K–12 schools, there is little research that has helped us understand this context and how it shapes STEM teachers' effect on PST learning at the subject matter or program level. Additional research within and across all of the various institutions that prepare PSTs would serve to deepen and expand the field's knowledge of how best to produce diverse and skilled groups of beginning teachers.

Methods

Methods vary with purposes, but the majority of studies in the field involves more qualitative study of local contexts. These studies, most often case studies, can yield rich descriptions but are not sufficient, in and of themselves, to support generalizations about STEM preparation and PSTs. Case studies, by definition, are not designed to support statistical generalization and, thus, need to be complemented with other methodological approaches.

Even studies that are more quantitative in nature, for the most part, also do not support generalizable findings because of limitations to sampling designs. This includes studies using surveys, observations, experiments, and quasi-experiments. There are a number of exceptions, including secondary analyses of state, national, and international data sets that were developed using representative sampling strategies.

Consistent with previous reviews (e.g., Cochran-Smith & Zeichner, 2013), the majority of methods in this review is characterized by researchers studying their own institutions and, often, their own courses. While this more participatory research can have unique value, it also brings with it certain limitations that are not present when research is conducted by those more distanced from the target of research.

In the studies reviewed here, researchers typically develop their own measures for studying phenomena of interest. This has a number of consequences for limiting our understanding of STEM teacher preparation. First, there is often little information available about the quality of the measures and the validity of inferences that researchers are making based on their data. Studies, for the most part, provide very little, if any, evidence about the accuracy, reliability, or validity of their measures. Not having such information limits the strength of findings of any particular study.

Second, the lack of common measures and the lack of evidence about locally developed measures also mean that studies may not always be exploring the same construct even if, nominally, it appears that they are. For example, while many studies explore PCK, because they operationalize PCK differently through their measures, the studies are often exploring very different conceptions of PCK. These variations must be accounted for when trying to synthesize information across studies. This lack of commonly accepted measures also makes it more difficult to compare results of studies that vary on key dimensions such as content, grade levels, and other contextual issues.

Stronger connections between studies are possible when they make use of common methods and measures, and this can help users understand the degrees to which findings in one context apply in other contexts and whether interventions related to the use of research evidence (URE) can be used broadly. We suggest that a more systematic approach for the sharing of research instruments, analytic schemes, operationalization of core components, and other tools be made available in one or more repositories focused on STEM teacher education research. This is not a simple task and will require thinking about the nature of metadata to organize such a repository and thinking about how to engage the field in contributing to, supporting, and using these repositories. Nevertheless, methodological repositories now exist in other fields (e.g., the health sciences) to develop synergies across research initiatives (see https://www.protocols.io/).

Recommendations for Future Research and Development Efforts

Considering the current status of research together with the social and organizational context previously described, we make two sets of recommendations regarding the research and development work that might benefit STEM teacher preparation moving forward. The first concerns the building of programs of research over time. The second involves a broadening of the object of inquiry.

Stepping back to look across the purposes of teacher preparation research, it is clear that there are few studies that belong to a reasonably well-defined program of scholarship that includes a full range of methodological approaches and integration of theory and empirical findings over time. We have seen some areas in education where headway has been made through this type of integration (e.g., teacher evaluation, early reading). The large number of descriptive and interpretive studies or studies that rely on limited samples is not, in itself, problematic. These studies have important value—they illuminate process and mechanisms, provide insights, and nominate critical factors. But by themselves they are limited in helping us build a comprehensive body of research to inform critical research and policy questions around STEM teacher preparation. Therefore, one set of suggestions for improving research and development in the preparation of STEM teachers for high-needs contexts concerns the deliberate building of lines of scholarship that include the varied ways of knowing critical to robust research understandings (Moss & Haertel, 2016).

Such research would, first and foremost, *track and report on foundational design information*—the race and gender of research participants; the accuracy and reliability of measures; as well as the subject matter and grade levels under investigation. We cannot know whether findings are robust if we do not have information about the reliability of the measures underlying those findings. Likewise, we cannot know whether grade level or subject matter is important in preparation if we do not track it carefully over a body of research.

Deliberate lines of scholarship would *make clear use of descriptive scholarship and appropriate samples* as that line of scholarship evolves and builds insights. Even in this limited review, there was a great deal of descriptive research that could and should serve as the basis for the development of measures as well as increasingly sophisticated and detailed theory. Such theory, measures, and insights can then undergird larger studies built on purposeful samples that allow for broader generalizations through, for example, quasi- and experimentally designed studies.

Without *common measures* it is very hard to build bodies of research that rely on comparable insights. Common measures are central to the improvement of STEM teacher preparation. Future research and development efforts should take up broad valued outcomes (i.e., not only the achievement of K–12 students or PSTs' scores on licensure exams) and pursue broad agreement on new measures that can be used across studies and in similar contexts. It is both inefficient and unhelpful to have the majority of researchers building new measures every time they carry out a new study. Certainly, measures should align to the intervention under study (Bell, Wilson, Higgins, & McCoach, 2010), but, where possible, include additional measures that have validity arguments in similar contexts and purposes that would build the field's knowledge base about both the new and more established measures and the constructs and processes of interest.

A second set of recommendations concerns the objects of inquiry, or what is being studied. The social and organizational context of teacher preparation suggests that researchers must treat the whole teacher preparation system as an object of inquiry. Taking just subject-matter knowledge as a case in point, we know little beyond small-scale descriptive work about the mathematics and science knowledge elementary teachers have prior to entry into preparation programs, as compared to the curriculum they will teach. It may be the case that everything elementary teachers learn about the mathematics relevant to the K–12 curriculum is learned outside of their preparation programs. Perhaps it is learned both prior to and inside of preparation programs. We do not know. Further, there is a great deal of scholarship on the predictive validity of teacher tests (e.g., Goldhaber, Gratz, & Theobald, 2017; Goldhaber, Krieg, & Theobold, 2014), but this tells us little about the content itself—what exactly PSTs know, who knows it, what institutions develop that knowledge, and how we might intervene to improve it before those PSTs begin their preparation programs. Careful studies of elementary and secondary teachers' STEM knowledge prior to their preparation programs would provide new insights and avenues to improve preparation. By broadening the lens of what counts as teacher preparation to include the IHE curriculum, the EPP-based teacher education curriculum, and teacher educators at IHEs and K–12 school placements, the field will be able to develop a more complete account of STEM teacher preparation.

Within teacher preparation programs, this review revealed there is uneven treatment of critical components of preparation and very little work beyond outcomes of teacher preparation programs at the program level. Future research and development work should productively develop theorized, synthetic accounts of what teacher preparation is, how it proceeds at a program level, what shapes the quality of a program, and those programs' social, emotional, cognitive, and practice outcomes. To summarize, we do not have an empirically grounded account of what is learned and how it is learned at a program level. We also do not have such accounts at the individual level. Whatever the reason, this state of research evidence is problematic if we hope to develop strong STEM teachers.

We must continue to develop our research on individuals' PST development; however, a straightforward analogy makes clear why existing research must be complemented with urgent program-level research. If we think about teacher preparation programs as if they were K–12 schools, the lack of focus at the program (school) level becomes stark. There are large bodies of research that consider how K–12 schools are organized; how they are improved; how leaders shape the outcomes and processes at those schools; and the wide range of outcomes—from social-emotional to academic—that K–12 students have. We do not have these bodies of research for teacher preparation programs. We have many studies that look at a single course—the parallel would be large numbers of studies of one K–12 teacher's classroom. While important and insightful, such studies do not help us understand how schools and classroom experiences within schools matter. Except for the conceptually thin metrics of teacher licensure tests, we have few common conceptions, tasks, or measures that define what we want PSTs to know and be able to do. Just as in the education of K–12 students, it is critical for teacher educators to decide on the important outcomes they hold for PSTs and then develop and use common tasks and measures to document what PSTs are and are not learning. Aside from two self-efficacy measures, this review did not turn up any measures used across multiple studies.

It will not be simple to develop the types of synthetic accounts that the field needs. In fact, such accounts require collaborations within and across EPPs in order to have large enough samples and multiple institutions for the types of generalization required. But such collaborations are necessary to begin to develop the knowledge base necessary for improving STEM preparation for diverse learners. Such work will not be easy. There is a long history and strong social norms against opening up teacher educators' practices to inspection and empirical scholarship. But there are new approaches, such as improvement science efforts happening at the Carnegie Foundation for the Advancement of Teaching (see https://www.carnegiefoundation.org/our-ideas/six-core-principles-improvement/) and some precedent within teacher education (e.g., the Core Practices Consortium [see https://www.corepracticeconsortium.com/about]; Grossman, 2018) for doing this work. If the field can learn from this work and focus on the development of new research and development agendas in such productive directions, research on STEM teacher preparation can rapidly improve.

Table 1

Number of Articles Reviewed, by Journal

Journal Title	Articles Reviewed
American Educational Research Journal	3
Educational Evaluation and Policy Analysis	4
Educational Policy	3
Educational Researcher	6
The Elementary School Journal	2
International Journal of Science and Mathematics Education	8
Journal for Research in Mathematics Education	2
Journal of Mathematics Teacher Education	31
Journal of Research in Science Teaching	4
Journal of Science Teacher Education	29
Journal of Teacher Education	34
Mathematical Thinking and Learning	0
Science Education	7
Teaching and Teacher Education	41

Table 2

Overview of Summarization Categories for Articles

Category	Sub-category	Description(s)
Purposes	Understanding STEM PST Learning and Development	These studies are primarily concerned with documenting the learning and development process of PSTs; examining PST learning outcomes in the areas of knowledge, practices, and/or beliefs; or studying interventions at the course, classroom, or instructor level.
	Improving EPPs	These studies refer to research that examines or evaluates interventions that are designed to make program improvements. The focus of these studies is at the program level, in contrast to Purpose1, which is focused on courses, classrooms, or instructors.
	Contributing to EPP Accountability	These studies have the purpose of adding evidence and methodological insights to research and policy debates focused on accountability issues in pre-service teacher preparation.
	Describing and Understanding Relationships Between STEM Teacher Preparation and Other Valued Outcomes	These studies focus on the relationships between pre-service teacher preparation and in-service teacher and student processes and outcomes, including but not limited to student achievement and teacher labor market decisions.
	Understanding Assessments and Measurement of STEM Teacher Preparation Quality	These studies focus on the validity and reliability of measures of teacher preparation quality, including value-added estimates and teacher performance assessments.

	Framing or Reframing Issues of STEM Teacher Preparation	This literature includes discussions of historical trends, reviews of literature, and constructions and/or critiques of conceptual frameworks, all used to consider different dimensions of teacher preparation.
	Understanding STEM Teacher Educators and Their Practices	These studies focus on teacher educators as individuals as well as how they learn to carry out the work of teaching PSTs. Research examines teacher educators' knowledge, practices, capabilities, beliefs, and identities as characteristics important to the learning opportunities provided to PSTs.
Researchers' Role		The role of the researcher vis a vis the study goals, study participants, study context, is this the researchers' home institution, and research-practice partnership
Sample	Geographic Location	Where does the study take place? (e.g., state, region [if more than one state])
	Sample Size	Identify the unit of analysis and then indicate the *n* for that unit. If there are multiple levels (e.g., 200 teachers nested in three programs), indicate each one.
	Grade Levels	The grade levels for which PSTs are preparing or the focal grades of study (K–12)
	Subject Matter	Science, technology, engineering, arts, mathematics, English language arts (ELA), social studies, physical education, other
	Racial Background	Racial backgrounds of PSTs
	Sampling Approach	Volunteer, convenience, stratified, etc. Be sure to note the population from which the sample is drawn. If there are multiple levels, please note the sampling approach for each level.
Methods and Methodological	Experimental or Quasi-experimental	Randomized trials, experiments with controls, and any analyses that are attempting to establish causal relationships using statistical design methods

Traditions	Survey Research	Includes large-scale surveys (e.g., SASS, TIMSS, NAEP) and more local surveys (e.g., preparation program, state, etc.)
	Interviews	Involve structured questioning of participants in some aspect of teacher preparation
	Design-based Research	Involves the design of interventions, studies of the design and ongoing improvements, and studies of the interdependencies of instructional design and theory development
	Ethnographic	In-depth qualitative studies of a particular educational context (e.g., first-hand experience and exploration of a particular social or cultural experience)
	Mixed Methods	Studies in which different and complementary methods are used to address the problem of interest
	Case Study	Comprehensive analysis (often using multiple methodologies and evidence sources) of a particular case; can also compare small numbers of cases (e.g., teacher education accreditation policies in three states)
	Critical Race Theory (or critical stances)	Uses a critical perspective to recognize and describe power and subjugation; often is skeptical about objectivity, meritocracy, etc.; frequently uses contextual/historical analysis and descriptions of the experiences of people of color
	Participatory Action Research	Characterized by goals and relationships between researcher and community in which research is carried out using many different methods; tends to be activist-oriented, with a focus on empowerment
	Literature Review	Articles that systematically review the literature on STEM teacher preparation
	Assessment Research	Describes and investigates the design, quality, and validity of measures of teacher preparation
	Conceptual Syntheses and Frameworks	Scholarly contributions that provide new insights into the frameworks and concepts used in research on teacher preparation

Phenomena of Inquiry		This is the phenomenon the researchers are trying to understand (e.g., student teaching experiences, teacher preparation curriculum, field placements, licensure, demographics, knowledge, beliefs, dispositions, program effectiveness, program accreditation).
Findings		Summary of medium- to high-level findings
Validity	Reliability and Accuracy	Accuracy and consistency of scores across relevant sources of variation (e.g., raters, testing occasions, time)
	Fairness	Evidence that explains the degree to which scores have the same meaning for all test takers and are not substantially influenced by construct-irrelevant barriers to individuals' performance
	Developer	By instrument, note if the developer is the researcher using the instrument.

[a] Classification by method was largely driven by statements the authors made about their selected methodology. However, there was a small number of cases in which the research team disagreed with the authors' claim about methodology and, therefore, assigned a different methodology to the study.

References

Adams, J. D., & Gupta, P. (2017). Informal science institutions and learning to teach: An examination of identity, agency, and affordances. *Journal of Research in Science Teaching*, *54*(1), 121–138. doi: 10.1002/tea.21270

Akerson, V. L., Khemmawadee, P., Park Rogers, M. A., Carter, I., & Galindo, E. (2017). Exploring the use of lesson study to develop elementary preservice teachers' pedagogical content knowledge for teaching nature of science. *International Journal of Science and Mathematics Education*, *15*(2), 293–312. doi: 10.1007/s10763-015-9690-x

Amador, J. (2016). Professional noticing practices of novice mathematics teacher educators. *International Journal of Science and Mathematics Education, 14*(1), 217–241. doi: 10.1007/s10763-014-9570-9

Amador, J. M., & Carter, I. S. (2018). Audible conversational affordances and constraints of verbalizing professional noticing during prospective teacher lesson study. *Journal of Mathematics Teacher Education*, *21*(1), 5–34. doi: 10.1007/s10857-016-9347-x

American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, *44*(8), 448–452. doi: 10.3102/0013189X15618385

American Educational Research Association (AERA), American Psychological Association (APA), & National Council of Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational

Research Association. Retrieved from http://www.aera.net/Publications/Books/Standards-for-Educational-Psychological-Testing-2014-Edition

Andrà, C., Brunetto, D., Levenson, E., & Liljedahl, P. (Eds.). (2017). *Teaching and learning in maths classrooms: Emerging themes in affect-related research: Teachers' beliefs, students' engagement and social interaction*. Cham, Switzerland: Springer International Publishing AG.

Avraamidou, L. (2014). Tracing a beginning elementary teacher’s development of identity for science teaching. *Journal of Teacher Education*, *65*(3), 223–240. doi: 10.1177/0022487113519476

Baeten, M., & Simons, M. (2014). Student teachers’ team teaching: Models, effects, and conditions for implementation. *Teaching and Teacher Education*, *41*, 92–110. doi: 10.1016/j.tate.2014.03.010

Bahr, D. L., Monroe, E. E., & Eggett, D. (2014). Structural and conceptual interweaving of mathematics methods coursework and field practica. *Journal of Mathematics Teacher Education*, *17*(3), 271–297. doi: 10.1007/s10857-013-9258-z

Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, *60*(5), 497–511. doi: 10.1177/0022487109348479

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content Knowledge for Teaching: What Makes It Special? Journal of Teacher Education, 59(5), 389-407. doi:10.1177/0022487108324554

Barnett, E., & Friedrichsen, P. J. (2015). Educative mentoring: How a mentor supported a preservice biology teacher's pedagogical content knowledge development. *Journal of Science Teacher Education*, *26*(7), 647–668. doi: 10.1007/s10972-015-9442-3

Bastian, K. C., & Henry, G. T. (2015). Teachers without borders: Consequences of teacher labor force mobility. *Educational Evaluation and Policy Analysis, 37*(2), 163–183. doi: 10.3102/0162373714535662

Bastian, K. C., Henry, G. T., Pan, Y., & Lys, D. (2016). Teacher candidate performance assessments: Local scoring and implications for teacher preparation program improvement. *Teaching and Teacher Education, 59,* 1–12. doi: 10.1016/j.tate.2016.05.008

Bates, L. A., & Glick, J. E. (2013). Does it matter if teachers and schools match the student? Racial and ethnic disparities in problem behaviors. *Social Science Research*, *42*(5), 1180–1190. doi: 10.1016/j.ssresearch.2013.04.005

Bautista, N. U., & Boone, W. J. (2015). Exploring the impact of TeachME™ lab virtual classroom teaching simulation on early childhood education majors' self-efficacy beliefs. *Journal of Science Teacher Education*, *26*(3), 237–262. doi: 10.1007/s10972-014-9418-8

Beilstein, S. O., Perry, M., & Bates, M. S. (2017). Prompting meaningful analysis from pre-service teachers using elementary mathematics video vignettes. *Teaching and Teacher Education*, *63*, 285–295. doi: 10.1016/j.tate.2017.01.005

Bell, C. A., Qi, Y., Croft, A. J., Leusner, D., Gitomer, D. H., & Pianta, R. C. (2013). Improving observational score quality: Challenges in observer thinking. In T. J. Kane, K. A. Kerr, &

R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project*. San Francisco, CA: Jossey-Bass. doi: 10.1002/9781119210856

Bell, C. A., Wilson, S. M., Higgins, T., & McCoach, D. B. (2010). Measuring the effects of professional development: The case of Developing Mathematical Ideas. *Journal for Research in Mathematics Education*, *41*(5), 479–512. doi: 10.2307/41110411

Benedict-Chambers, A., & Aram, R. (2017). Tools for teacher noticing: Helping preservice teachers notice and analyze student thinking and scientific practice use. *Journal of Science Teacher Education*, *28*(3), 294–318. doi: 10.1080/1046560X.2017.1302730

Bill and Melinda Gates Foundation. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Author. Retrieved from http://k12education.gatesfoundation.org/resource/gathering-feedback-on-teaching-combining-high-quality-observations-with-student-surveys-and-achievement-gains-3/

Bleiler, S. K. (2015). Increasing awareness of practice through interaction across communities: The lived experiences of a mathematician and mathematics teacher educator. *Journal of Mathematics Teacher Education*, *18*(3), 231–252. doi: 10.1007/s10857-014-9275-6

Blumenreich, M., & Gupta, A. (2015). The globalization of Teach for America: An analysis of the institutional discourses of Teach for America and Teach for India within local contexts. *Teaching and Teacher Education*, *48*, 87–96. doi: 10.1016/j.tate.2015.01.017

Borko, H., Stecher, B., & Kuffner, K. (2007). *Using artifacts to characterize reform-oriented instruction: The Scoop Notebook and rating guide* (CSE Technical Report 707). Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing (CRESST)/UCLA. Retrieved from cresst.org/wp-content/uploads/R707.pdf

Bottoms, S. I., Ciechanowski, K., Jones, K., de la Hoz, J., & Fonseca, A. L. (2017). Leveraging the community context of Family Math and Science Nights to develop culturally responsive teaching practices. *Teaching and Teacher Education, 61,* 1–15. doi: 10.1016/j.tate.2016.09.006

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31*(4), 416–440. doi: 10.3102/0162373709353129

Bravo, M. A., Mosqueda, E., Solís, J. L., & Stoddart, T. (2014). Possibilities and limits of integrating science and diversity education in preservice elementary teacher preparation. *Journal of Science Teacher Education*, *25*(5), 601–619. doi: 10.1007/s10972-013-9374-8

Brown, C. P., Englehardt, J., & Mathers, H. (2016). Examining preservice teachers' conceptual and practical understandings of adopting iPads into their teaching of young children. *Teaching and Teacher Education*, *60*, 179–190. doi: 10.1016/j.tate.2016.08.018

Bryant, C. L., Maarouf, S., Burcham, J., & Greer, D. (2016). The examination of a teacher candidate assessment rubric: A confirmatory factor analysis. *Teaching and Teacher Education*, *57,* 79–96. doi: 10.1016/j.tate.2016.03.012

Bullough, R. V. (2014). Toward reconstructing the narrative of teacher education. *Journal of Teacher Education*, *65*(3), 185–194. doi: 10.1177/0022487113519131

Bybee, R. W. (2014). NGSS and the next generation of science teachers. *Journal of Science Teacher Education, 25*(2), 211–221. doi: 10.1007/s10972-014-9381-4

Cappella, E., Aber, J. L., & Kim, H. Y. (2016). Teaching beyond achievement tests: Perspectives from developmental and education science. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 249–347). Washington, DC: American Educational Research Association.

Casabianca, J. M., Lockwood, J. R., & McCaffrey, D. F. (2014). Trends in classroom observation scores. *Educational and Psychological Measurement*, *75*(2), 311–337. doi: 10.1177/0013164414539163

Castro Superfine, A., & Li, W. (2014). Exploring the mathematical knowledge needed for teaching teachers. *Journal of Teacher Education*, *65*(4), 303–314. doi: 10.1177/0022487114534265

Caughlan, S., & Heng, J. (2014). Observation and teacher quality: Critical analysis of observational instruments in preservice teacher performance assessment. *Journal of Teacher Education*, *65*(5), 375–388. doi: 10.1177/0022487114541546

Chazan, D., Herbst, P. G., & Clark, L. M. (2016). Research on the teaching of mathematics: A call to theorize the role of society and schooling in mathematics instruction. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 1039–

1097). Washington, DC: American Educational Research Association. doi: 10.3102/978-0-935302-48-6_17

Christ, T., Arya, P., & Chiu, M. M. (2017). Video use in teacher education: An international survey of practices. *Teaching and Teacher Education*, *63*, 22–35. doi: 10.1016/j.tate.2016.12.005

Clark, S. K., Byrnes, D., & Sudweeks, R. R. (2015). A comparative examination of student teacher and intern perceptions of teaching ability at the preservice and inservice stages. *Journal of Teacher Education*, *66*(2), 170–183. doi: 10.1177/0022487114561659

Clotfelter, C., Ladd, H. F., & Vigdor, J. (2007). *Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects* (NBER Working Paper 13617). Cambridge, Ma: National Bureau of Economic Research. Retrieved from https://econpapers.repec.org/paper/nbrnberwo/13617.htm

Cochran-Smith, M. (2016). Teaching and teacher education: Absence and presence in AERA presidential addresses. *Educational Researcher*, *45*(2), 92–99. doi: 10.3102/0013189X16639040

Cochran-Smith, M., Ell, F., Grudnoff, L., Haigh, M., Hill, M., & Ludlow, L. (2016). Initial teacher education: What does it take to put equity at the center? *Teaching and Teacher Education*, *57*, 67–78. doi: 10.1016/j.tate.2016.03.006

Cochran-Smith, M., Stern, R., Sánchez, J. G., Miller, A., Keefe, E. S., Fernández, M. B., . . . Baker, M. (2016). *Holding teacher preparation accountable: A review of claims and* *evidence*. Boulder, CO: National Education Policy Center. Retrieved from http://nepc.colorado.edu/publication/teacher-prep

Cochran-Smith, M., Villegas, A. M., Abrams, L. W., Chavez-Moreno, L. C., Mills, T., & Stern, R. (2016). Research on teacher preparation: Charting the landscape of a sprawling field. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 439–547). Washington, DC: American Educational Research Association. doi: 10.3102/978-0-935302-48-6_7

Cochran-Smith, M., & Zeichner, K. M. (Eds.) (2013). *Studying teacher education: The report on the AERA panel on research and teacher education*. New York, NY: Routledge.

Corbin, J., & Strauss, A. (2008). Book review: Basics of qualitative research: Techniques and procedures for developing grounded theory (3rd ed.). *Organizational Research Methods, 12*(3), 614–617. doi: 10.1177/1094428108324514

Council for the Accreditation of Educator Preparation. (2016a). *2013 CAEP standards*. Washington, DC: Author. Retrieved from http://caepnet.org/standards/introduction

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.

Danielson, C. (2011). *The Framework for Teaching evaluation instrument*. Princeton, NJ: The Danielson Group.

Darling-Hammond, L. (2016). Research on teaching and teacher education and its influences on policy and practice. *Educational Researcher, 45*(2), 83–91. doi: 10.3102/0013189X16639597

Darling-Hammond, L., & Bransford, J. (Eds.). (2007). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA: Jossey-Bass.

Darling-Hammond, L., & Hyler, M. (2013). The role of performance assessment in developing teaching as a profession. *Rethinking Schools, 27*(4), 1–5. Retrieved from https://www.rethinkingschools.org/articles/the-role-of-performance-assessment-in-developing-teaching-as-a-profession

Davis, E. A., & Boerst, T. A. (2014). *Designing elementary teacher education to prepare well-started beginners* (TeachingWorks Working Paper). Ann Arbor, MI: TeachingWorks, University of Michigan School of Education. Retrieved from http://www.teachingworks.org/research-data/workingpapers

Davis, E. A., Kloser, M., Wells, A., Windschitl, M., Carlson, J., & Marino, J.-C. (2017). Teaching the practice of leading sense-making discussions in science: Science teacher educators using rehearsals. *Journal of Science Teacher Education, 28*(3), 275–293. doi: 10.1080/1046560X.2017.1302729

Diezmann, C. M., & Watters, J. J. (2015). The knowledge base of subject matter experts in teaching: A case study of a professional scientist as beginning teacher. *International Journal of Science and Mathematics Education*, *13*(6), 1517–1537. doi: 10.1007/s10763-014-9561-x

Donna, J. D., & Hick, S. R. (2017). Developing elementary preservice teacher subject matter knowledge through the use of educative science curriculum materials. *Journal of Science Teacher Education*, *28*(1), 92–110. doi: 10.1080/1046560X.2017.1279510

Dotger, B. H. (2015). Core pedagogy. *Journal of Teacher Education*, *66*(3), 215–226. doi: 10.1177/0022487115570093

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review, 45*, 44–52. doi: 0.1016/j.econedurev.2015.01.007

Eisenhart, M. (2001). Changing conceptions of culture and ethnographic methodology: Recent thematic shifts and their implications for research on teaching. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 209–225). Washington, DC: American Educational Research Association.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119–161). New York, NY: MacMillan.

Evans, C. M. (2017). Predictive validity and impact of CAEP standard 3.2: Results from one Master's-level teacher preparation program. *Journal of Teacher Education*, *68*(4), 363–376. doi: 10.1177/0022487117702577

Felton-Koestler, M. D. (2017). Mathematics education as sociopolitical: Prospective teachers' views of the what, who, and how. *Journal of Mathematics Teacher Education*, *20*(1), 49–74. doi: 10.1007/s10857-015-9315-x

Forzani, F. M. (2014). Understanding "core practices" and "practice-based" teacher education. *Journal of Teacher Education*, *65*(4), 357–368. doi: 10.1177/0022487114533800

Fuller, E. J. (2014). Shaky methods, shaky motives: A critique of the National Council of Teacher Quality's review of teacher preparation programs. *Journal of Teacher Education, 65*(1), 63–77. doi: 10.1177/0022487113503872

Gareis, C. R., & Grant, L. W. (2014). The efficacy of training cooperating teachers. *Teaching and Teacher Education*, *39*, 77–88. doi: 10.1016/j.tate.2013.12.007

George, A. L., & Bennett, A. (2004). *Case studies and theory development in the social sciences*. Cambridge, MA: The MIT Press.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review, 52*, 209–224. doi: 10.1016/j.econedurev.2016.03.002

Ghousseini, H. (2015). Core practices and problems of practice in learning to lead classroom discussions. *The Elementary School Journal*, *115*(3), 334–357. doi: 10.1086/680053

Ghousseini, H., & Herbst, P. (2016). Pedagogies of practice and opportunities to learn about classroom mathematics discussions. *Journal of Mathematics Teacher Education*, *19*(1), 79–103. doi: 10.1007/s10857-014-9296-1

Gitomer, D. H., Phelps, G., Weren, B., Howell, H. & Croft, A. J. (2014). Evidence on the validity of content knowledge for teaching assessments. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project* (pp. 493–528). San Francisco, CA: Jossey-Bass.

Gitomer, D. H., & Zisk, R. C. (2015). Knowing what teachers know. *Review of Research in Education, 39*(1), 1–53. doi: 10.3102/0091732X14557001

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Publishing Company.

Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *The Journal of Human Resources, 42*(4), 765–794. doi: 10.3368/jhr.XLII.4.765

Goldhaber, D., Cowan, J., & Theobald, R. (2017). Evaluating prospective teachers: Testing the predictive validity of the edTPA. *Journal of Teacher Education, 68*(4), 377–393. doi:10.1177/0022487117702582

Goldhaber, D., Gratz, T., & Theobald, R. (2017). What's in a teacher test? Assessing the relationship between teacher licensure test scores and student STEM achievement and course-taking. *Economics of Education Review, 61*, 112–129. doi:10.1016/j.econedurev.2017.09.002

Goldhaber, D., Krieg, J., & Theobald, R. (2014). Knocking on the door to the teaching profession? Modeling the entry of prospective teachers into the workforce. *Economics of Education Review, 43*, 106–124. doi: 10.1016/j.econedurev.2014.10.003

Goldhaber, D., Krieg, J. M., & Theobald, R. (2017). Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness. *American Educational Research Journal*, *54*(2), 325–359. doi: 10.3102/0002831217690516

Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review*, *34*, 29–44. doi: 10.1016/j.econedurev.2013.01.011

Goodwin, A. L., Smith, L., Souto-Manning, M., Cheruvu, R., Tan, M. Y., Reed, R., & Taveras, L. (2014). What should teacher educators know and be able to do? Perspectives from practicing teacher educators. *Journal of Teacher Education*, *65*(4), 284–302. doi: 10.1177/0022487114535266

Grossman, P. (Ed.). (2018). *Teaching core practices in teacher education*. (Core Practices in Education Series). Cambridge, MA: Harvard Education Press.

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching*, *15*(2), 273–289. doi: 10.1080/13540600902875340

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal*, *45*(1), 184–205. doi: 10.3102/0002831207312906

Hansen, M., Backes, B., & Brady, V. (2016). Teacher attrition and mobility during the Teach for America clustering strategy in Miami-Dade County Public Schools. *Educational Evaluation and Policy Analysis*, *38*(3), 495–516. doi: 10.3102/0162373716638441

Herman, B. C., & Clough, M. P. (2016). Teachers' longitudinal NOS understanding after having completed a science teacher education program. *International Journal of Science and Mathematics Education*, *14*(Suppl 1), S207–S227. doi: 10.1007/s10763-014-9594-1

Hiebert, J., Miller, E., & Berk, D. (2017). Relationships between mathematics teacher preparation and graduates' analyses of classroom teaching. *The Elementary School Journal*, *117*(4), 687–707. doi: 10.1086/691685

Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. Journal for Research in Mathematics Education, 41(5), 513-545.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*(4), 430–511. doi: 10.1080/07370000802177235

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406. doi: 10.3102/00028312042002371

Hjalmarson, M. A. (2017). Learning to teach mathematics specialists in a synchronous online course: A self-study. *Journal of Mathematics Teacher Education*, *20*(3), 281–301. doi: 10.1007/s10857-015-9323-x

Hoffman, J. V., Wetzel, M. M., Maloch, B., Greeter, E., Taylor, L., DeJulio, S., & Vlach, S. K. (2015). What can we learn from studying the coaching interactions between cooperating teachers and preservice teachers? A literature review. *Teaching and Teacher Education*, *52*, 99–112. doi: 10.1016/j.tate.2015.09.004

Hohensee, C. (2017). Preparing elementary prospective teachers to teach early algebra. *Journal of Mathematics Teacher Education*, *20*(3), 231–257. doi: 10.1007/s10857-015-9324-9

Horizon Research, Inc. (2000). *Inside the classroom: Observation and analytic protocol*. Chapel Hill, NC: Author. Retrieved from www.horizon-research.com/instruments/clas/cop.pdf

Horm, D. M., Hyson, M., & Winton, P. J. (2013). Research on early childhood teacher education: Evidence from three domains and recommendations for moving forward. *Journal of Early Childhood Teacher Education*, *34*(1), 95-112. doi:10.1080/10901027.2013.758541

Ingvarson, L., & Rowley, G. (2017). Quality assurance in teacher education and outcomes: A study of 17 countries. *Educational Researcher*, *46*(4), 177–193. doi: 10.3102/0013189X17711900

Jacobson, E. D. (2017). Field experience and prospective teachers' mathematical knowledge and beliefs. *Journal for Research in Mathematics Education*, *48*(2), 148–190. doi: 10.5951/jresematheduc.48.2.0148

Janssen, F., Grossman, P., & Westbroek, H. (2015). Facilitating decomposition and recomposition in practice-based teacher education: The power of modularity. *Teaching and Teacher Education*, *51*, 137–146. doi: 10.1016/j.tate.2015.06.009

Johnson, H. J., & Cotterman, M. E. (2015). Developing preservice teachers' knowledge of science teaching through video clubs. *Journal of Science Teacher Education*, *26*(4), 393–417. doi: 10.1007/s10972-015-9429-0

Jong, C., & Hodges, T. E. (2015). Assessing attitudes toward mathematics across teacher education contexts. *Journal of Mathematics Teacher Education*, *18*(5), 407–425. doi: 10.1007/s10857-015-9319-6

Kahn, S., & Zeidler, D. L. (2016). Using our heads and HARTSS*: Developing perspective-taking skills for socioscientific reasoning (*humanities, ARTs, and social sciences). *Journal of Science Teacher Education*, *27*(3), 261–281. doi: 10.1007/s10972-016-9458-3

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). New York, NY: American Council on Education, Macmillan Publishing.

Kang, H., & Anderson, C. W. (2015). Supporting preservice science teachers' ability to attend and respond to student thinking by design. *Science Education*, *99*(5), 863–895. doi: 10.1002/sce.21182

Katsh-Singer, R., McNeill, K. L., & Loper, S. (2016). Scientific argumentation for all? Comparing teacher beliefs about argumentation in high, mid and low SES schools. *Science Education*, *100*(3), 410–436. doi: 10.1002/sce.21214

Kazemi, E., Ghousseini, H., Cunard, A., & Turrou, A. C. (2016). Getting inside rehearsals: Insights from teacher educators to support work on complex practice. *Journal of Teacher Education*, *67*(1), 18–31. doi: 10.1177/0022487115615191

Kersting, N. B., Givvin, K. B., Thompson, B., Santagata, R., & Stigler, J. (2012). Developing measures of usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, *49*(3), 568–590. doi: 10.3102/0002831212437853

Kesner, J., Kwon, K.-A., & Lim, C. (2016). The role of race in preservice teachers' perceptions of and attitudes towards corporal punishment & child maltreatment. *Teaching and Teacher Education*, *59*, 318–326. doi: 10.1016/j.tate.2016.06.015

Kim, E., & Corcoran, R. P. (2017). How engaged are pre-service teachers in the United States? *Teaching and Teacher Education*, *66*, 12–23. doi: 10.1016/j.tate.2017.03.020

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2015). Teacher preparation programs and teacher quality: Are there real differences across programs? *Education Finance and Policy, 10*(4), 508–534. doi: 10.1162/EDFP_a_00172

Krieg, J. M., Theobald, R., & Goldhaber, D. (2016). A foot in the door: Exploring the role of student teaching assignments in teachers' initial job placements. *Educational Evaluation and Policy Analysis*, *38*(2), 364–388. doi: 10.3102/0162373716630739

Labaree, D. F. (2006). *The trouble with ed schools*. New Haven, CT: Yale University Press.

Ladson-Billings, G., & Tate, W. F. (1995). Toward a critical race theory of education. *Teachers College Record*, *97*(1), 47–68.

Lahey, T. (2017). Collaborating to address the challenge of academic language. *Journal of Teacher Education, 68*(3), 239–250. doi: 10.1177/0022487117696279

Lannin, J. K., Webb, M., Chval, K., Arbaugh, F., Hicks, S., Taylor, C., & Bruton, R. (2013). The development of beginning mathematics teacher pedagogical content knowledge. *Journal of Mathematics Teacher Education*, *16*(6), 403–426. doi: 10.1007/s10857-013-9244-5

Lasley, T. J., Siedentop, D., & Yinger, R. (2006). A systemic approach to enhancing teacher quality: The Ohio model. *Journal of Teacher Education*, *57*(1), 13–21. doi: 10.1177/0022487105284455

Levine, A. (2006). *Educating school teachers*. Washington, DC: The Education Schools Project. Retrieved from edschools.org/pdf/Educating_Teachers_Report.pdf

Li, W., & Castro Superfine, A. (2018). Mathematics teacher educators' perspectives on their design of content courses for elementary preservice teachers. *Journal of Mathematics Teacher Education*, *21*(2), 179–201. doi: 10.1007/s10857-016-9356-9

Lincove, J. A., Osborne, C., Mills, N., & Bellows, L. (2015). Teacher preparation for profit or prestige: Analysis of a diverse market for teacher preparation. *Journal of Teacher Education*, *66*(5), 415–434. doi: 10.1177/0022487115602311

Liu, S., Bell, C. A., Jones, N., & McCaffrey, D. F. (2019). Classroom observation systems in context: A case for the validation of observation systems. *Educational Assessment, Evaluation and Accountability*. doi: 10.1007/s11092-018-09291-3

Lovin, L. H., Stevens, A. L., Siegfried, J., Wilkins, J. L. M., & Norton, A. (2018). Pre-K-8 prospective teachers' understanding of fractions: An extension of fractions schemes and operations research. *Journal of Mathematics Teacher Education*, *21*(3), 207–235. doi: 10.1007/s10857-016-9357-8

Lynch, H. L., & Fisher-Ari, T. R. (2017). Metaphor as pedagogy in teacher education. *Teaching and Teacher Education*, *66*, 195–203.doi: 10.1016/j.tate.2017.03.021

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. New York, NY: Routledge.

Mansour, N. (2009). Science teachers' beliefs and practices: Issues, implications and research agenda. *International Journal of Environmental and Science Education*, *4*(1), 25–48. Retrieved from https://eric.ed.gov/?id=EJ884384

Marinell, W. H., & Johnson, S. M. (2014). Midcareer entrants to teaching: Who they are and how they may, or may not, change teaching. *Educational Policy*, *28*(6), 743–779. doi: 10.1177/0895904813475709

Martínez, J. F., Borko, H., Stecher, B. M., Luskin, R., & Kloser, M. (2012). Measuring classroom assessment practice using instructional artifacts: A validation study of the QAS Notebook. *Educational Assessment, 17*(2-3), 107–131. doi: 10.1080/10627197.2012.715513

Maxwell, B., Tremblay-Laprise, A.-A., Filion, M., Boon, H., Daly, C., van den Hoven, M., . . . Walters, S. (2016). A five-country survey on ethics education in preservice teaching programs. *Journal of Teacher Education*, *67*(2), 135–151. doi: 10.1177/0022487115624490

McNew-Birren, J., & van den Kieboom, L. A. (2017). Exploring the development of core teaching practices in the context of inquiry-based science instruction: An interpretive case study. *Teaching and Teacher Education*, *66,* 74–87. doi: 10.1016/j.tate.2017.04.001

Menon, D., & Sadler, T. D. (2016). Preservice elementary teachers' science self-efficacy beliefs and science content knowledge. *Journal of Science Teacher Education*, *27*(6), 649–673. doi: 10.1007/s10972-016-9479-y

Menon, D., & Sadler, T. D. (2018). Sources of science teaching self-efficacy for preservice elementary teachers in science content courses. *International Journal of Science and Mathematics Education*, *16*(5), 835–855. doi: 10.1007/s10763-017-9813-7

Mihaly, K., & McCaffrey, D. F. (2014). Grade level variation in observational measures of teacher effectiveness. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project* (pp. 9–49). San Francisco, CA: Jossey-Bass.

Mitchell, R. M., & Marin, K. A. (2015). Examining the use of a structured analysis framework to support prospective teacher noticing. *Journal of Mathematics Teacher Education*, *18*(6), 551–575. doi: 10.1007/s10857-014-9294-3

Moore-Russo, D. A., & Wilsey, J. N. (2014). Delving into the meaning of productive reflection: A study of future teachers' reflections on representations of teaching. *Teaching and Teacher Education*, *37*, 76–90. doi: 10.1016/j.tate.2013.10.002

Morris, A. K., & Hiebert, J. (2017). Effects of teacher preparation courses: Do graduates use what they learned to plan mathematics lessons? *American Educational Research Journal*, *54*(3), 524–567. doi: 10.3102/0002831217695217

Moss, P. A., & Haertel, E. H. (2016). Engaging methodological pluralism. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 127–147). Washington, DC: American Educational Research Association.

National Council on Teacher Quality. (2013). *Teacher prep review: A review of the nation's teacher preparation programs*. Washington, DC: Author. Retrieved from http://www.nctq.org/dmsView/Teacher_Prep_Review_2013_Report

National Research Council. (2000). *Educating teachers of science, mathematics, and technology: New practices for the new millennium*. Washington, DC: The National Academies Press. doi: 10.17226/9832

National Research Council. (2002). *Scientific research in education*. Washington, DC: The National Academies Press. doi: 10.17226/10236.

National Research Council. (2005). *Advancing scientific research in education*. Washington, DC: The National Academies Press. doi: /10.17226/11112

National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Washington, DC: The National Academies Press. doi: 10.17226/12882

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press. doi: 10.17226/13165

Olson, J. K., Bruxvoort, C. N., & Vande Haar, A. J. (2016). The impact of video case content on preservice elementary teachers’ decision-making and conceptions of effective science teaching. *Journal of Research in Science Teaching, 53*(10), 1500–1523. doi: 10.1002/tea.21335

Özçınar, H. (2015). Mapping teacher education domain: A document co-citation analysis from 1992 to 2012. *Teaching and Teacher Education*, *47*, 42–61. doi: 10.1016/j.tate.2014.12.006

Pappamihiel, N. E., Ousley-Exum, D., & Ritzhaupt, A. (2017). The impact of digital stories on preservice teacher beliefs about English language learners. *Teaching and Teacher Education*, *67*, 171–178. doi: 10.1016/j.tate.2017.06.014

Paufler, N. A., & Amrein-Beardsley, A. (2016). Preparing teachers for educational renewal within current contexts of accountability. *Journal of Teacher Education*, *67*(4), 251–262. doi: 10.1177/0022487116660154

Pianta, R. C., Hamre, B. K., Haynes, N. J., Mintz, S. L., & La Paro, K. M. (2005). *Classroom Assessment Scoring System (CLASS), secondary manual*. Charlottesville, VA: University of Virginia Center for Advanced Study of Teaching and Learning

Preston, C. (2017). University-based teacher preparation and middle grades teacher effectiveness. *Journal of Teacher Education*, *68*(1), 102–116. doi: 10.1177/0022487116660151

Qian, H., & Youngs, P. (2016). The effect of teacher education programs on future elementary mathematics teachers' knowledge: A five-country analysis using TEDS-M data. *Journal of Mathematics Teacher Education*, *19*(4), 371–396. doi: 10.1007/s10857-014-9297-0

Reagan, E. M., Chen, C., & Vernikoff, L. (2016). "Teachers are works in progress": A mixed methods study of teaching residents' beliefs and articulations of teaching for social justice. *Teaching and Teacher Education*, *59*, 213–227. doi: 10.1016/j.tate.2016.05.011

Redding, C., & Smith, T. M. (2016). Easy in, easy out: Are alternatively certified teachers turning over at increased rates? *American Educational Research Journal*, *53*(4), 1086–1125. doi: 10.3102/0002831216653206

Reeves, T. D., & Honig, S. L. (2015). A classroom data literacy intervention for pre-service teachers. *Teaching and Teacher Education*, *50*, 90–101. doi: 10.1016/j.tate.2015.05.007

Riegle-Crumb, C., Morton, K., Moore, C., Chimonidou, A., Labrake, C., & Kopp, S. (2015). Do inquiring minds have positive attitudes? The science education of preservice elementary teachers. *Science Education*, *99*(5), 819–836. doi: 10.1002/sce.21177

Roller, S. A. (2016). What they notice in video: A study of prospective secondary mathematics teachers learning to teach. *Journal of Mathematics Teacher Education*, *19*(5), 477–498. doi: 10.1007/s10857-015-9307-x

Ronfeldt, M. (2015). Field placement schools and instructional effectiveness. *Journal of Teacher Education*, *66*(4), 304–320. doi: 10.1177/0022487115592463

Ronfeldt, M., & Campbell, S. L. (2016). Evaluating teacher preparation using graduates' observational ratings. *Educational Evaluation and Policy Analysis*, *38*(4), 603–625. doi: 10.3102/0162373716649690

Ronfeldt, M., & Reininger, M. (2012). More or better student teaching? *Teaching and Teacher Education*, *28*(8), 1091–1106. doi: 10.1016/j.tate.2012.06.003

Ronfeldt, M., Reininger, M., & Kwok, A. (2013). Recruitment or preparation? Investigating the effects of teacher characteristics and student teaching. *Journal of Teacher Education*, *64*(4), 319–337. doi: 10.1177/0022487113488143

Rowan, B., & Raudenbush, S. W. (2016). Teacher evaluation in American schools. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 1159–1216). Washington, DC: American Educational Research Association.

Rubin, B. C., Abu El-Haj, T. R., Graham, E., & Clay, K. (2016). Confronting the urban civic opportunity gap: Integrating youth participatory action research into teacher education. *Journal of Teacher Education*, *67*(5), 424–436. doi: 10.1177/0022487116667195

Sabel, J. L., Forbes, C. T., & Zangori, L. (2015). Promoting prospective elementary teachers' learning to use formative assessment for life science instruction. *Journal of Science Teacher Education*, *26*(4), 419–445. doi: 10.1007/s10972-015-9431-6

Saçkes, M., & Trundle, K. C. (2014). Preservice early childhood teachers' learning of science in a methods course: Examining the predictive ability of an intentional learning model. *Journal of Science Teacher Education*, *25*(4), 413–444. doi: 10.1007/s10972-013-9355-y

Santagata, R., & Yeh, C. (2014). Learning to teach mathematics and to analyze teaching effectiveness: Evidence from a video- and practice-based approach. *Journal of Mathematics Teacher Education, 17*(6), 491–514. doi: 10.1007/s10857-013-9263-2

Santau, A. O., Maerten-Rivera, J. L., Bovis, S., & Orend, J. (2014). A mile wide or an inch deep? Improving elementary preservice teachers' science content knowledge within the context of a science methods course. *Journal of Science Teacher Education*, *25*(8), 953–976. doi: 0.1007/s10972-014-9402-3

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton, Mifflin and Company.

Shane, J. W., Binns, I. C., Meadows, L., Hermann, R. S., & Benus, M. J. (2016). Beyond evolution: Addressing broad interactions between science and religion in science teacher education. *Journal of Science Teacher Education*, *27*(2), 165–181. doi: 10.1007/s10972-016-9449-4

Shaw, J. M., Lyon, E. G., Stoddart, T., Mosqueda, E., & Menon, P. (2014). Improving science and literacy learning for English language learners: Evidence from a pre-service teacher preparation intervention. *Journal of Science Teacher Education*, *25*(5), 621–643. doi: 10.1007/s10972-013-9376-6

Shuls, J. V., & Trivitt, J. R. (2015). Teacher effectiveness: An analysis of licensure screens. *Educational Policy*, *29*(4), 645–675. doi: 10.1177/0895904813510777

Sleeter, C. (2014). Toward teacher education research that informs policy. *Educational Researcher*, *43*(3), 146–153. doi: 10.3102/0013189X14528752

Soslau, E. (2015). Development of a post-lesson observation conferencing protocol: Situated in theory, research, and practice. *Teaching and Teacher Education*, *49*, 22–35. doi: 10.1016/j.tate.2015.02.012

Steele, A. (2016). Troubling STEM: Making a case for an ethics/STEM partnership. *Journal of Science Teacher Education, 27*(4), 357–371. doi: 10.1007/s10972-016-9463-6

Subramaniam, K. (2014). Prospective secondary mathematics teachers’ pedagogical knowledge for teaching the estimation of length measurements. *Journal of Mathematics Teacher Education*, *17*(2), 177–198. doi: 10.1007/s10857-013-9255-2

Sun, J., & van Es, E. A. (2015). An exploratory study of the influence that analyzing teaching has on preservice teachers' classroom practice. *Journal of Teacher Education, 66*(3), 201–214. doi: 10.1177/0022487115574103

Swanson, L. H., & Coddington, L. R. (2016). Creating partnerships between teachers & undergraduates interested in secondary math & science education. *Teaching and Teacher Education, 59*, 285–294. doi: 10.1016/j.tate.2016.06.008

Swars, S. L., Smith, S. Z., Smith, M. E., Carothers, J., & Myers, K. (2018). The preparation experiences of elementary mathematics specialists: Examining influences on beliefs, content knowledge, and teaching practices. *Journal of Mathematics Teacher Education, 21*(2), 123–145. doi: 10.1007/s10857-016-9354-y

Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Rowley, G., Peck, R., . . . Reckase, M. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. East Lansing, MI: Teacher Education International Study Center, College of Education, Michigan State University.

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202–248. doi: 10.3102/00346543068002202

Thanheiser, E. (2015). Developing prospective teachers' conceptions with well-designed tasks: Explaining successes and analyzing conceptual difficulties. *Journal of Mathematics Teacher Education*, *18*(2), 141–172. doi: 10.1007/s10857-014-9272-9

Thomson, M. M., & Palermo, C. (2014). Preservice teachers' understanding of their professional goals: Case studies from three different typologies. *Teaching and Teacher Education*, *44*, 56–68. doi: 10.1016/j.tate.2014.08.002

Trauth-Nare, A. (2015). Influence of an intensive, field-based life science course on preservice teachers' self-efficacy for environmental science teaching. *Journal of Science Teacher Education*, *26*(5), 497–519. doi: 10.1007/s10972-015-9434-3

Tuck, E., & Gorlewski, J. (2016). Racist ordering, settler colonialism, and edTPA: A participatory policy analysis. *Educational Policy*, *30*(1), 197–217. doi: 10.1177/0895904815616483

Trygstad, P. J., Smith, P. S., Banilower, E. R., & Nelson, M. M. (2013). *The status of elementary science education: Are we ready for the next generation science standards?* Chapel Hill, NC: Horizon Research, Inc..

Tygret, J. A. (2017). The influence of student teachers on student achievement: A case study of teacher perspectives. *Teaching and Teacher Education*, *66*, 117–126. doi: 10.1016/j.tate.2017.04.005

Tyminski, A. M., Zambak, V. S., Drake, C., & Land, T. J. (2014). Using representations, decomposition, and approximations of practices to support prospective elementary mathematics teachers' practice of organizing discussions. *Journal of Mathematics Teacher Education*, *17*(5), 463–487. doi: 10.1007/s10857-013-9261-4

Weiland, I. S., Hudson, R. A., & Amador, J. M. (2014). Preservice formative assessment interviews: The development of competent questioning. *International Journal of Science and Mathematics Education*, *12*(2), 329–352. doi: 10.1007/s10763-013-9402-3

Wenzel, S., Nagaoka, J. K., Morris, L., Billings, S., & Fendt, C. (2002). *Documentation of the 1996–2002 Chicago Annenberg Research Project strand on authentic intellectual demand exhibited in assignments and student work: A technical process manual*. Chicago, IL: Consortium on Chicago School Research. Retrieved from http://ccsr.uchicago.edu/sites/default/files/publications/p67.pdf

Whitacre, I., & Nickerson, S. D. (2016). Prospective elementary teachers making sense of multidigit multiplication: Leveraging resources. *Journal for Research in Mathematics Education*, *47*(3), 270–307. doi: 10.5951/jresematheduc.47.3.0270

Whitehurst, G. J., Chingos, M. M., & Lindquist, K. M. (2014). *Evaluating teachers with classroom observations: Lessons learned in four districts*. Washington, DC: Brown Center on Education Policy at Brookings. Retrieved from https://www.brookings.edu/research/evaluating-teachers-with-classroom-observations-lessons-learned-in-four-districts/

Wilson, S. (2011). *Effective STEM teacher preparation, induction, and professional development*. Paper commissioned for the Workshop on Successful STEM Education in K–12 Schools, convened by the Board on Science Education with support from the National Science Foundation. Washington, DC: The National Academies of Sciences, Engineering, and Medicine. Retrieved from http://sites.nationalacademies.org/dbasse/bose/dbasse_080128#.UgEMEFPkDDn

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2001). Teacher preparation research: Current knowledge, gaps, and recommendations. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington. Retrieved from https://www.education.uw.edu/ctp/content/teacher-preparation-research-current-knowledge-gaps-and-recommendations

Windschitl, M., & Barton, A. C. (2016). Rigor and equity by design: Locating a set of core teaching practices for the science education community. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 1099–1158). Washington, DC: American Educational Research Association. doi: 10.3102/978-0-935302-48-6_18

Windschitl, M. A., & Stroupe, D. (2017). The three-story challenge: Implications of the Next Generation Science Standards for teacher preparation. *Journal of Teacher Education, 68*(3), 251–261. doi: 10.1177/0022487117696278

Yeh, C., & Santagata, R. (2015). Preservice teachers' learning to generate evidence-based hypotheses about the impact of mathematics teaching on learning. *Journal of Teacher Education*, *66*(1), 21–34. doi: 10.1177/0022487114549470

Zeichner, K. M. (2013). A research agenda for teacher education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report on the AERA panel on research and teacher education* (pp. 737–759). New York, NY: Routledge.

Zeichner, K., Payne, K. A., & Brayko, K. (2015). Democratizing teacher education. *Journal of Teacher Education, 66*(2), 122–135. doi: 10.1177/0022487114560908

Appendix

All Articles Summarized and Reviewed

For each article we asked two questions: 1) Is the population of study participants exclusively STEM teachers? and 2) Are the findings framed such that there are direct implications for STEM teacher education? Articles were counted as focusing on STEM as long as there was at least one yes response.

*Focused on STEM.

AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. (2015). *Educational Researcher*, *44*(8), 448–452. doi: 10.3102/0013189X15618385

*Adams, J. D., & Gupta, P. (2017). Informal science institutions and learning to teach: An examination of identity, agency, and affordances. *Journal of Research in Science Teaching, 54*(1), 121–138. doi: 10.1002/tea.21270

*Akerson, V. L., Khemmawadee, P., Park Rogers, M. A.,Carter, I., & Galindo, E. (2017). Exploring the use of lesson study to develop elementary preservice teachers' pedagogical content knowledge for teaching nature of science. *International Journal of Science and Mathematics Education*, *15*(2), 293–312. doi: 10.1007/s10763-015-9690-x

*Amador, J. (2016). Professional noticing practices of novice mathematics teacher educators. *International Journal of Science and Mathematics Education, 14*(1), 217–241. doi: 10.1007/s10763-014-9570-9

*Amador, J. M., & Carter, I. S. (2018). Audible conversational affordances and constraints of verbalizing professional noticing during prospective teacher lesson study. *Journal of Mathematics Teacher Education*, *21*(1), 5–34. doi: 10.1007/s10857-016-9347-x

*Anhalt, C. O., & Cortez, R. (2016). Developing understanding of mathematical modeling in secondary teacher preparation. *Journal of Mathematics Teacher Education, 19*(6), 523–

545. doi: 10.1007/s10857-015-9309-8

*Avraamidou, L. (2014). Developing a reform-minded science teaching identity: The role of informal science environments. *Journal of Science Teacher Education, 25*(7), 823–843. doi: 10.1007/s10972-014-9395-y

Baeten, M., & Simons, M. (2014). Student teachers' team teaching: Models, effects, and conditions for implementation. *Teaching and Teacher Education, 41*, 92–110. doi: 10.1016/j.tate.2014.03.010

*Bahr, D. L., Monroe, E. E., & Eggett, D. (2014). Structural and conceptual interweaving of mathematics methods coursework and field practica. *Journal of Mathematics Teacher Education, 17*(3), 271–297. doi: 10.1007/s10857-013-9258-z

*Barnett, E., & Friedrichsen, P. J. (2015). Educative mentoring: How a mentor supported a preservice biology teacher's pedagogical content knowledge development. *Journal of Science Teacher Education, 26*(7), 647–668. doi: 10.1007/s10972-015-9442-3

*Barnhart, T., & van Es, E. (2015). Studying teacher noticing: Examining the relationship among pre-service science teachers' ability to attend, analyze and respond to student thinking. *Teaching and Teacher Education, 45*, 83–93. doi: 10.1016/j.tate.2014.09.005

Bastian, K. C., & Henry, G. T. (2015). Teachers without borders: Consequences of teacher labor force mobility. *Educational Evaluation and Policy Analysis, 37*(2), 163–183. doi: 10.3102/0162373714535662

Bastian, K. C., Henry, G. T., Pan, Y., Lys, D. (2016). Teacher candidate performance assessments: Local scoring and implications for teacher preparation program improvement. *Teaching and Teacher Education, 59*, 1–12. doi: 10.1016/j.tate.2016.05.008

*Bautista, N. U., & Boone, W. J. (2015). Exploring the impact of TeachME™ lab virtual classroom teaching simulation on early childhood education majors' self-efficacy beliefs. *Journal of Science Teacher Education, 26*(3), 237–262. doi: 10.1007/s10972-014-9418-8

*Beilstein, S. O., Perry, M., & Bates, M. S. (2017). Prompting meaningful analysis from pre-service teachers using elementary mathematics video vignettes. *Teaching and Teacher Education, 63,* 285–295. doi: 10.1016/j.tate.2017.01.005

*Benedict-Chambers, A., & Aram, R. (2017). Tools for teacher noticing: Helping preservice teachers notice and analyze student thinking and scientific practice use. *Journal of Science Teacher Education, 28*(3), 294–318. doi: 10.1080/1046560X.2017.1302730

*Bleiler, S. K. (2015). Increasing awareness of practice through interaction across communities: The lived experiences of a mathematician and mathematics teacher educator. *Journal of Mathematics Teacher Education, 18*(3), 231–252. doi: 10.1007/s10857-014-9275-6

Blumenreich, M., & Gupta, A. (2015). The globalization of Teach for America: An analysis of the institutional discourses of Teach for America and Teach for India within local contexts. *Teaching and Teacher Education, 48*, 87–96. doi: 10.1016/j.tate.2015.01.017

*Borgerding, L. A., Klein, V. A., Ghosh, R., & Eibel, A. (2015). Student teachers' approaches to teaching biological evolution. *Journal of Science Teacher Education, 26*(4), 371–392. doi: 10.1007/s10972-015-9428-1

*Bottoms, S. I., Ciechanowski, K., Jones, K., de la Hoz, J., & Fonseca, A. L. (2017). Leveraging the community context of Family Math and Science Nights to develop culturally responsive teaching practices. *Teaching and Teacher Education, 61*, 1–15. doi: 10.1016/j.tate.2016.09.006

Bowman, M., & Gottesman, I. (2017). Making the socio-historical visible: A place-conscious approach to social foundations in practice-centered teacher preparation. *Teaching and Teacher Education, 68*, 232–240. doi: 10.1016/j.tate.2017.09.001

*Bravo, M. A., Mosqueda, E., Solís, J. L., & Stoddart, T. (2014). Possibilities and limits of integrating science and diversity education in preservice elementary teacher preparation. *Journal of Science Teacher Education, 25*(5), 601–619. doi: 10.1007/s10972-013-9374-8

*Brown, C. P., Englehardt, J., & Mathers, H. (2016). Examining preservice teachers' conceptual and practical understandings of adopting iPads into their teaching of young children. *Teaching and Teacher Education, 60*, 179–190. doi: 10.1016/j.tate.2016.08.018

Bryant, C. L., Maarouf, S., Burcham, J., & Greer, D. (2016). The examination of a teacher candidate assessment rubric: A confirmatory factor analysis. *Teaching and Teacher Education, 57*, 79–96. doi: 10.1016/j.tate.2016.03.012

Bullough, R. V., Jr. (2014). Toward reconstructing the narrative of teacher education: A rhetorical analysis of preparing teachers. *Journal of Teacher Education, 65*(3), 185–194. doi: 10.1177/0022487113519131

*Carpenter, J. P., Tur, G., & Marín, V. I. (2016). What do U.S. and Spanish pre-service teachers think about educational and professional use of Twitter? A comparative study. *Teaching and Teacher Education, 60*, 131–143. doi: 10.1016/j.tate.2016.08.011

*Castro Superfine, A., & Li, W. (2014). Exploring the mathematical knowledge needed for teaching teachers. *Journal of Teacher Education, 65*(4), 303–314. doi: 10.1177/0022487114534265

Caughlan, S., & Heng, J. (2014). Observation and teacher quality: Critical analysis of observational instruments in preservice teacher performance assessment. *Journal of Teacher Education, 65*(5), 375–388. doi: 10.1177/0022487114541546

*Charalambous, C. Y. (2015). Working at the intersection of teacher knowledge, teacher beliefs, and teaching practice: A multiple-case study. *Journal of Mathematics Teacher Education, 18*(5), 427–445. doi: 10.1007/s10857-015-9318-7

Choi, E., Gaines, R. E., Park, J. H., Williams, K. M., Schallert, D. L., Yu, L., & Lee, J. (2016). Small stories in online classroom discussion as resources for preservice teachers' making sense of becoming a bilingual educator. *Teaching and Teacher Education*, *58*, 1–16. doi: 10.1016/j.tate.2016.03.015

Christ, T., Arya, P., & Chiu, M. M. (2017). Video use in teacher education: An international survey of practices. *Teaching and Teacher Education, 63,* 22–35. doi: 10.1016/j.tate.2016.12.005

*Cian, H., Dsouza, N., Lyons, R., & Cook, M. (2017). Influences on the development of inquiry-based practices among preservice teachers. *Journal of Science Teacher Education, 28*(2), 186–204. doi: 10.1080/1046560X.2016.1277832

Clark, S. K., Byrnes, D., & Sudweeks, R. R. (2015). A comparative examination of student teacher and intern perceptions of teaching ability at the preservice and inservice stages. *Journal of Teacher Education, 66*(2), 170–183. doi: 10.1177/0022487114561659

Cochran-Smith, M. (2016). Teaching and teacher education: Absence and presence in AERA presidential addresses. *Educational Researcher*, *45*(2), 92–99. doi: 10.3102/0013189X16639040

Cochran-Smith, M., Ell, F., Grudnoff, L., Haigh, M., Hill, M., & Ludlow, L. (2016). Initial teacher education: What does it take to put equity at the center? *Teaching and Teacher Education, 57,* 67–78. doi: 10.1016/j.tate.2016.03.006

Cochran-Smith, M., & Villegas, A. M. (2015). Framing teacher preparation research: An overview of the field, part 1. *Journal of Teacher Education, 66*(1), 7–20. doi:10.1177/0022487114549072

Cochran-Smith, M., Villegas, A. M., Abrams, L., Chavez-Moreno, L., Mills, T., & Stern, R. (2015). Critiquing teacher preparation research: An overview of the field, part II. *Journal of Teacher Education, 66*(2), 109–121. doi: 10.1177/0022487114558268

*Colwell, J., & Enderson, M. C. (2016). “When I hear literacy”: Using pre-service teachers' perceptions of mathematical literacy to inform program changes in teacher education. *Teaching and Teacher Education, 53*, 63–74. doi: 10.1016/j.tate.2015.11.001

Darling-Hammond, L. (2016). Research on teaching and teacher education and its influences on policy and practice. *Educational Researcher, 45*(2), 83–91. doi: 10.3102/0013189X16639597

*Davis, E. A., Kloser, M., Wells, A., Windschitl, M., Carlson, J., & Marino, J.-C. (2017). Teaching the practice of leading sense-making discussions in science: Science teacher educators using rehearsals. *Journal of Science Teacher Education, 28*(3), 275–293. doi: 10.1080/1046560X.2017.1302729

*Diezmann, C. M., & Watters, J. J. (2015). The knowledge base of subject matter experts in teaching: A case study of a professional scientist as beginning teacher. *International Journal of Science and Mathematics Education, 13*(6), 1517–1537. doi: 10.1007/s10763-014-9561-x

*Donna, J. D., & Hick, S. R. (2017). Developing elementary preservice teacher subject matter knowledge through the use of educative science curriculum materials. *Journal of Science Teacher Education, 28*(1), 92–110. doi: 10.1080/1046560X.2017.1279510

Dotger, B. H. (2015). Core pedagogy: Individual uncertainty, shared practice, formative ethos.

Journal of Teacher Education, 66(3), 215–226. doi: 10.1177/0022487115570093

*Dotger, B., Masingila, J., Bearkland, M., & Dotger, S. (2015). Exploring iconic interpretation and mathematics teacher development through clinical simulations. *Journal of Mathematics Teacher Education, 18*(6), 577–601. doi: 10.1007/s10857-014-9290-7

Drake, C., Land, T. J., & Tyminski, A. M. (2014). Using educative curriculum materials to support the development of prospective teachers' knowledge. *Educational Researcher, 43*(3), 154–162. doi: 10.3102/0013189X14528039

*Earnest, D., & Amador, J. M. (2017). Lesson planimation: Prospective elementary teachers' interactions with mathematics curricula. *Journal of Mathematics Teacher Education*, 1–32. doi: 10.1007/s10857-017-9374-2

Evans, C. M. (2017). Predictive validity and impact of CAEP standard 3.2: Results from one Master's-level teacher preparation program. *Journal of Teacher Education, 68*(4), 363–376. doi: 10.1177/0022487117702577

*Felton-Koestler, M. D. (2017). Mathematics education as sociopolitical: Prospective teachers' views of the what, who, and how. *Journal of Mathematics Teacher Education, 20*(1), 49–74. doi: 10.1007/s10857-015-9315-x

Forzani, F. M. (2014). Understanding "core practices" and "practice-based" teacher education: Learning from the past. *Journal of Teacher Education, 65*(4), 357–368. doi: 10.1177/0022487114533800

Fuller, E. J. (2014). Shaky methods, shaky motives: A critique of the National Council of Teacher Quality's review of teacher preparation programs. *Journal of Teacher Education, 65*(1), 63–77. doi: 10.1177/0022487113503872

Gareis, C. R., & Grant, L. W. (2014). The efficacy of training cooperating teachers. *Teaching* *and Teacher Education, 39*, 77–88. doi: 10.1016/j.tate.2013.12.007

Gelfuso, A., & Dennis, D. V. (2014). Getting reflection off the page: The challenges of developing support structures for pre-service teacher reflection. *Teaching and Teacher Education, 38*, 1–11. doi: 10.1016/j.tate.2013.10.012

*Ghousseini, H. (2015). Core practices and problems of practice in learning to lead classroom discussions. *The Elementary School Journal, 115*(3), 334–357. doi: 10.1086/680053

*Ghousseini, H., & Herbst, P. (2016). Pedagogies of practice and opportunities to learn about classroom mathematics discussions. *Journal of Mathematics Teacher Education, 19*(1), 79–103. doi: 10.1007/s10857-014-9296-1

*Glaze, A. L., Goldston, M. J., & Dantzler, J. (2015). Evolution in the southeastern USA: Factors influencing acceptance and rejection in pre-service science teachers. *International Journal of Science and Mathematics Education, 13*(6), 1189–1209. doi: 10.1007/s10763-014-9541-1

*Goldhaber, D., Krieg, J. M., & Theobald, R. (2017). Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness. *American Educational Research Journal, 54*(2), 325–359. doi: 10.3102/0002831217690516

Goodwin, A. L., Smith, L., Souto-Manning, M., Cheruvu, R., Tan, M. Y., Reed, R., & Taveras, L. (2014). What should teacher educators know and be able to do? Perspectives from practicing teacher educators. *Journal of Teacher Education, 65*(4), 284–302. doi: 10.1177/0022487114535266

*Gunckel, K. L., & Wood, M. B. (2016). The principle-practical discourse edge: Elementary preservice and mentor teachers working together on colearning tasks. *Science Education, 100*(1), 96–121. doi: 10.1002/sce.21187

Hansen, M., Backes, B., & Brady, V. (2016). Teacher attrition and mobility during the Teach for America clustering strategy in Miami-Dade County Public Schools. *Educational Evaluation and Policy Analysis, 38*(3), 495–516. doi: 10.3102/0162373716638441

*Hanuscin, D. L., & Zangori, L. (2016). Developing practical knowledge of the Next Generation Science Standards in elementary science teacher education. *Journal of Science Teacher Education, 27*(8), 799–818. doi: 10.1007/s10972-016-9489-9

*Harlow, D. B., Swanson, L. H., & Otero, V. K. (2014). Prospective elementary teachers' analysis of children's science talk in an undergraduate physics course *Journal of Science Teacher Education, 25*(1), 97–117. doi: 10.1007/s10972-012-9319-7

*Hawkins, S., & Park Rogers, M. (2016). Tools for reflection: Video-based reflection within a preservice community of practice. *Journal of Science Teacher Education, 27*(4), 415–437. doi: 10.1007/s10972-016-9468-1

*Herman, B. C., & Clough, M. P. (2016). Teachers' longitudinal NOS understanding after having completed a science teacher education program. *International Journal of Science and Mathematics Education, 14*(Suppl 1), S207–S227. doi: 10.1007/s10763-014-9594-1

*Hernandez, C., & Shroyer, M. G. (2017). The use of culturally responsive teaching strategies among Latina/o student teaching interns during science and mathematics instruction of CLD students. *Journal of Science Teacher Education, 28*(4), 367–387. doi: 10.1080/1046560X.2017.1343605

*Hiebert, J., Miller, E., & Berk, D. (2017). Relationships between mathematics teacher preparation and graduates' analyses of classroom teaching. *The Elementary School Journal, 117*(4), 687–707. doi: 10.1086/691685

*Hjalmarson, M. A. (2017). Learning to teach mathematics specialists in a synchronous online course: A self-study. *Journal of Mathematics Teacher Education, 20*(3), 281–301. doi: 10.1007/s10857-015-9323-x

*Hodges, T. E., & Hodge, L. L. (2017). Unpacking personal identities for teaching mathematics within the context of prospective teacher education. *Journal of Mathematics Teacher Education, 20*(2), 101–118. doi: 10.1007/s10857-015-9339-2

Hoffman, J. V., Wetzel, M. M., Maloch, B., Greeter, E., Taylor, L., DeJulio, S., & Vlach, S. K. (2015). What can we learn from studying the coaching interactions between cooperating teachers and preservice teachers? A literature review. *Teaching and Teacher Education, 52*, 99–112. doi: 10.1016/j.tate.2015.09.004

*Hohensee, C. (2017). Preparing elementary prospective teachers to teach early algebra. *Journal of Mathematics Teacher Education, 20*(3), 231–257. doi: 10.1007/s10857-015-9324-9

*Hutner, T. L., & Markman, A. B. (2017). Applying a goal-driven model of science teacher cognition to the resolution of two anomalies in research on the relationship between science teacher education and classroom practice. *Journal of Research in Science Teaching, 54*(6), 713–736. doi: 10.1002/tea.21383

Ingvarson, L., & Rowley, G. (2017). Quality assurance in teacher education and outcomes: A study of 17 countries. *Educational Researcher, 46*(4), 177–193. doi: 10.3102/0013189X17711900

*Jacobson, E. D. (2017). Field experience and prospective teachers' mathematical knowledge and beliefs. *Journal for Research in Mathematics Education, 48*(2), 148–190. doi: 10.5951/jresematheduc.48.2.0148

Janssen, F., Grossman, P., & Westbroek, H. (2015). Facilitating decomposition and recomposition in practice-based teacher education: The power of modularity. *Teaching* *and Teacher Education, 51*, 137–146. doi: 10.1016/j.tate.2015.06.009

*Johnson, H. J., & Cotterman, M. (2015). Developing preservice teachers' knowledge of science teaching through video clubs. *Journal of Science Teacher Education*, *26*(4), 393–417. doi: 10.1007/s10972-015-9429-0

*Jong, C., & Hodges, T. E. (2015). Assessing attitudes toward mathematics across teacher education contexts. *Journal of Mathematics Teacher Education, 18*(5), 407–425. doi: 10.1007/s10857-015-9319-6

Kahn, S., & Zeidler, D. L. (2016). Using our heads and HARTSS: Developing perspective-taking skills for socioscientific reasoning (*humanities, ARTs, and social sciences). *Journal of Science Teacher Education, 27*(3), 261–281. doi: 10.1007/s10972-016-9458-3

*Kang, H., & Anderson, C. W. (2015). Supporting preservice science teachers' ability to attend and respond to student thinking by design. *Science Education, 99*(5), 863–895. doi: 10.1002/sce.21182

*Kazemi, E., Ghousseini, H., Cunard, A., & Turrou, A. C. (2016). Getting inside rehearsals: Insights from teacher educators to support work on complex practice. *Journal of Teacher Education, 67*(1), 18–31. doi: 10.1177/0022487115615191

Kesner, J., Kwon, K.-A., & Lim, C. (2016). The role of race in preservice teachers’ perceptions of and attitudes towards corporal punishment & child maltreatment. *Teaching and Teacher Education*, *59*, 318–326. doi: 10.1016/j.tate.2016.06.015

*Kier, M. W., & Lee, T. D. (2017). Exploring the role of identity in elementary preservice teachers who plan to specialize in science teaching. *Teaching and Teacher Education*, *61*, 199–210. doi: 10.1016/j.tate.2016.10.016

Kim, E., & Corcoran, R. P. (2017). How engaged are pre-service teachers in the United States?

Teaching and Teacher Education, 66, 12–23. doi: 10.1016/j.tate.2017.03.020

Krieg, J. M., Theobald, R., & Goldhaber, D. (2016). A foot in the door: Exploring the role of student teaching assignments in teachers' initial job placements. *Educational Evaluation and Policy Analysis, 38*(2), 364–388. doi: 10.3102/0162373716630739

Krutka, D. G., Bergman, D. J., Flores, R., Mason, K., & Jack, A. R. (2014). Microblogging about teaching: Nurturing participatory cultures through collaborative online reflection with pre-service teachers. *Teaching and Teacher Education, 40*, 83–93. doi: 10.1016/j.tate.2014.02.002

Lahey, T. (2017). Collaborating to address the challenge of academic language. *Journal of Teacher Education, 68*(3), 239–250. doi: 10.1177/0022487117696279

*Land, T. J., Tyminski, A. M., Drake, C. (2015). Examining pre-service elementary mathematics teachers' reading of educative curriculum materials. *Teaching and Teacher Education, 51*, 16–26. doi: 10.1016/j.tate.2015.05.009

*Lannin, J. K., Webb, M., Chval, K., Arbaugh, F., Hicks, S., Taylor, C., & Bruton, R. (2013). The development of beginning mathematics teacher pedagogical content knowledge. *Journal of Mathematics Teacher Education, 16*(6), 403–426. doi: 10.1007/s10857-013-9244-5

*Larkin, D. B., & Perry-Ryder, G. M. (2015). Without the light of evolution: A case study of resistance and avoidance in learning to teach high school biology. *Science Education, 99*(3), 549–576. doi: 10.1002/sce.21149

*Lee, J. (2016). Implementing college and career standards in math methods course for early childhood and elementary education teacher candidates. *International Journal of Science and Mathematics Education, 14*(1), 177–192. doi: 10.1007/s10763-014-9551-z

Lee, S., & Schallert, D. L. (2016). Becoming a teacher: Coordinating past, present, and future selves with perspectival understandings about teaching. *Teaching and Teacher Education, 56*, 72–83. doi: 10.1016/j.tate.2016.02.004

*Li, W., & Castro Superfine, A. (2018). Mathematics teacher educators' perspectives on their design of content courses for elementary preservice teachers. *Journal of Mathematics Teacher Education, 21*(2), 179–201. doi: 10.1007/s10857-016-9356-9

Lincove, J. A., Osborne, C., Mills, N., & Bellows, L. (2015). Teacher preparation for profit or prestige: Analysis of a diverse market for teacher preparation. *Journal of Teacher Education, 66*(5), 415–434. doi: 10.1177/0022487115602311

Loughran, J. (2014). Professionally developing as a teacher educator. *Journal of Teacher Education, 65*(4), 271–283. doi: 10.1177/0022487114533386

*Lovett, J. N., & Lee, H. S. (2017). New standards require teaching more statistics: Are preservice secondary mathematics teachers ready? *Journal of Teacher Education, 68*(3), 299–311. doi: 10.1177/0022487117697918

*Lovin, L. H., Stevens, A. L., Siegfried, J., Wilkins, J. L. M., & Norton, A. (2018). Pre-K–8 prospective teachers' understanding of fractions: An extension of fractions schemes and operations research. *Journal of Mathematics Teacher Education, 21*(3), 207–235. doi: 10.1007/s10857-016-9357-8

*Lui, A. M., & Bonner, S. M. (2016). Preservice and inservice teachers' knowledge, beliefs, and instructional planning in primary school mathematics. *Teaching and Teacher Education, 56*, 1–13. doi: 10.1016/j.tate.2016.01.015

Lynch, H. L., & Fisher-Ari, T. R. (2017). Metaphor as pedagogy in teacher education. *Teaching and Teacher Education, 66*, 195–203. doi: 10.1016/j.tate.2017.03.021

Madden, B. (2015). Pedagogical pathways for Indigenous education with/in teacher education. *Teaching and Teacher Education, 51*, 1–15. doi: 10.1016/j.tate.2015.05.005

Mansfield, C. F., Beltman, S., Broadley, T., & Weatherby-Fell, N. (2016). Building resilience in teacher education: An evidenced informed framework. *Teaching and Teacher Education, 54*, 77–87. doi: 10.1016/j.tate.2015.11.016

Marinell, W. H., & Johnson, S. M. (2014). Midcareer entrants to teaching: Who they are and how they may, or may not, change teaching. *Educational Policy, 28*(6), 743–779. doi: 10.1177/0895904813475709

Matias, C. E., & Grosland, T. J. (2016). Digital storytelling as racial justice: Digital hopes for deconstructing whiteness in teacher education. *Journal of Teacher Education, 67*(2), 152–164. doi: 10.1177/0022487115624493

Matsko, K. K., & Hammerness, K. (2014). Unpacking the "urban" in urban teacher education: Making a case for context-specific preparation. *Journal of Teacher Education, 65*(2), 128–144. doi: 10.1177/0022487113511645

Maxwell, B., Tremblay-Laprise, A.-A., Filion, M., Boon, H., Daly, C., van den Hoven, M., . . . Walters, S. (2016). A five-country survey on ethics education in preservice teaching programs. *Journal of Teacher Education, 67*(2), 135–151. doi: 10.1177/0022487115624490

*McNew-Birren, J., & van den Kieboom, L. A. (2017). Exploring the development of core teaching practices in the context of inquiry-based science instruction: An interpretive case study. *Teaching and Teacher Education, 66*, 74–87. doi: 10.1016/j.tate.2017.04.001

*Menon, D., & Sadler, T. D. (2016). Preservice elementary teachers' science self-efficacy beliefs and science content knowledge. *Journal of Science Teacher Education, 27*(6), 649–673.

doi: 10.1007/s10972-016-9479-y

*Menon, D., & Sadler, T. D. (2018). Sources of science teaching self-efficacy for preservice elementary teachers in science content courses. *International Journal of Science and Mathematics Education*, *16*(5), 835–855. doi: 10.1007/s10763-017-9813-7

Mills, C., & Ballantyne, J. (2016). Social justice and teacher education: A systematic review of empirical work in the field. *Journal of Teacher Education*, *67*(4), 263–276. doi: 10.1177/0022487116660152

*Mitchell, R. M., & Marin, K. A. (2015). Examining the use of a structured analysis framework to support prospective teacher noticing. *Journal of Mathematics Teacher Education, 18*(6), 551–575. doi: 10.1007/s10857-014-9294-3

*Moore-Russo, D. A., & Wilsey, J. N. (2014). Delving into the meaning of productive reflection: A study of future teachers' reflections on representations of teaching. *Teaching and Teacher Education, 37*, 76–90. doi: 10.1016/j.tate.2013.10.002

*Morris, A. K., & Hiebert, J. (2017). Effects of teacher preparation courses: Do graduates use what they learned to plan mathematics lessons? *American Educational Research Journal, 54*(3), 524–567. doi: 10.3102/0002831217695217

Moulding, L. R., Stewart, P. W., & Dunmeyer, M. L. (2014). Pre-service teachers' sense of efficacy: Relationship to academic ability, student teaching placement characteristics, and mentor support. *Teaching and Teacher Education, 41*, 60–66. doi: 10.1016/j.tate.2014.03.007

Naidoo, K., & Kirch, S. A. (2016). Candidates use a new teacher development process, transformative reflection, to identify and address teaching and learning problems in their work with children. *Journal of Teacher Education, 67*(5), 379–391. doi:

10.1177/0022487116653659

*Olson, J. K., Bruxvoort, C. N., & Vande Haar, A. J. (2016). The impact of video case content on preservice elementary teachers' decision-making and conceptions of effective science teaching. *Journal of Research in Science Teaching, 53*(10), 1500–1523. doi: 10.1002/tea.21335

*Olson, J. K., Tippett, C. D., Milford, T. M., Ohana, C., & Clough, M. P. (2015). Science teacher preparation in a North American context. *Journal of Science Teacher Education, 26*(1), 7–28. doi: 10.1007/s10972-014-9417-9

Özçınar, H. (2015). Mapping teacher education domain: A document co-citation analysis from 1992 to 2012. *Teaching and Teacher Education, 47*, 42–61. doi: 10.1016/j.tate.2014.12.006

Pappamihiel, N. E., Ousley-Exum, D., & Ritzhaupt, A. (2017). The impact of digital stories on preservice teacher beliefs about English language learners. *Teaching and Teacher Education, 67*, 171–178. doi: 10.1016/j.tate.2017.06.014

Paufler, N. A., & Amrein-Beardsley, A. (2016). Preparing teachers for educational renewal within current contexts of accountability: Reflecting upon John Goodlad's twenty postulates. *Journal of Teacher Education, 67*(4), 251–262. doi: 10.1177/0022487116660154

*Plummer, J. D., & Tanis Ozcelik, A. (2015). Preservice teachers developing coherent inquiry investigations in elementary astronomy. *Science Education, 99*(5), 932–957. doi: 10.1002/sce.21180

*Preston, C. (2017). University-based teacher preparation and middle grades teacher effectiveness. *Journal of Teacher Education, 68*(1), 102–116. doi:

10.1177/0022487116660151

*Qian, H., & Youngs, P. (2016). The effect of teacher education programs on future elementary mathematics teachers' knowledge: A five-country analysis using TEDS-M data. *Journal of Mathematics Teacher Education, 19*(4), 371–396. doi: 10.1007/s10857-014-9297-0

Reagan, E. M., Chen, C., & Vernikoff, L. (2016). “Teachers are works in progress”: A mixed methods study of teaching residents’ beliefs and articulations of teaching for social justice. *Teaching and Teacher Education*, *59*, 213–227. doi: 10.1016/j.tate.2016.05.011

Redding, C., & Smith, T. M. (2016). Easy in, easy out: Are alternatively certified teachers turning over at increased rates? *American Educational Research Journal, 53*(4), 1086–1125. doi: 10.3102/0002831216653206

Reeves, T. D. (2017). Pre-service teachers’ data use opportunities during student teaching. *Teaching and Teacher Education*, *63*, 263–273. doi: 10.1016/j.tate.2017.01.003

Reeves, T. D., & Honig, S. L. (2015). A classroom data literacy intervention for pre-service teachers. *Teaching and Teacher Education, 50*, 90–101. doi: 10.1016/j.tate.2015.05.007

*Richmond, G., Parker, J. M., & Kaldaras, L. (2016). Supporting reform-oriented secondary science teaching through the use of a framework to analyze construction of scientific explanations. *Journal of Science Teacher Education, 27*(5), 477–493. doi: 10.1007/s10972-016-9470-7

*Riegle-Crumb, C., Morton, K., Moore, C., Chimonidou, A., Labrake, C., & Kopp, S. (2015). Do inquiring minds have positive attitudes? The science education of preservice elementary teachers. *Science Education*, *99*(5), 819–836. doi: 10.1002/sce.21177

Riley, K., & Solic, K. (2017). "Change happens beyond the comfort zone": Bringing undergraduate teacher-candidates into activist teacher communities. *Journal of Teacher*

Education, *68*(2), 179–192. doi: 10.1177/0022487116687738

*Roller, S. A. (2016). What they notice in video: A study of prospective secondary mathematics teachers learning to teach. *Journal of Mathematics Teacher Education, 19*(5), 477–498. doi: 10.1007/s10857-015-9307-x

*Ronfeldt, M. (2015). Field placement schools and instructional effectiveness. *Journal of Teacher Education, 66*(4), 304–320. doi: 10.1177/0022487115592463

Ronfeldt, M., & Campbell, S. L. (2016). Evaluating teacher preparation using graduates' observational ratings. *Educational Evaluation and Policy Analysis, 38*(4), 603–625. doi: 10.3102/0162373716649690

Ronfeldt, M., & Reininger, M. (2012). More or better student teaching? *Teaching and Teacher Education*, *28*(8), 1091–1106. doi: 10.1016/j.tate.2012.06.003

Ronfeldt, M., Reininger, M., & Kwok, A. (2013). Recruitment or preparation? Investigating the effects of teacher characteristics and student teaching. *Journal of Teacher Education*, *64*(4), 319–337. doi: 10.1177/0022487113488143

*Ross, D. K., & Carter, J. L. (2015). Developing pre-service elementary teachers' pedagogical practices while planning using the learning cycle. *Journal of Science Teacher Education, 26*(6), 573–591. doi: 10.1007/s10972-015-9439-y

*Roth McDuffie, A., Foote, M. Q., Bolson, C., Turner, E. E., Aguirre, J. M., Bartell, T. G., . . . Land, T. (2014). Using video analysis to support prospective K–8 teachers' noticing of students' multiple mathematical knowledge bases. *Journal of Mathematics Teacher Education, 17*(3), 245–270. doi: 10.1007/s10857-013-9257-0

Rubin, B. C., Abu El-Haj, T. R., Graham, E., & Clay, K. (2016). Confronting the urban civic opportunity gap: Integrating youth participatory action research into teacher education.

Journal of Teacher Education, 67(5), 424–436. doi: 10.1177/0022487116667195

*Sabel, J. L., Forbes, C. T., & Zangori, L. (2015). Promoting prospective elementary teachers' learning to use formative assessment for life science instruction. *Journal of Science Teacher Education, 26*(4), 419–445. doi: 10.1007/s10972-015-9431-6

*Saçkes, M., & Trundle, K. C. (2014). Preservice early childhood teachers' learning of science in a methods course: Examining the predictive ability of an intentional learning model. *Journal of Science Teacher Education, 25*(4), 413–444. doi: 10.1007/s10972-013-9355-y

*Salter, I. Y., & Atkins, L. J. (2014). What students say versus what they do regarding scientific inquiry. *Science Education, 98*(1), 1–35. doi: 10.1002/sce.21084

*Santagata, R., & Yeh, C. (2014). Learning to teach mathematics and to analyze teaching effectiveness: Evidence from a video- and practice-based approach. *Journal of Mathematics Teacher Education, 17*(6), 491–514. doi: 10.1007/s10857-013-9263-2

*Santau, A. O., Maerten-Rivera, J. L., Bovis, S., & Orend, J. (2014). A mile wide or an inch deep? Improving elementary preservice teachers' science content knowledge within the context of a science methods course. *Journal of Science Teacher Education, 25*(8), 953–976. doi: 10.1007/s10972-014-9402-3

Schultz, K., & Ravitch, S. M. (2013). Narratives of learning to teach: Taking on professional identities. *Journal of Teacher Education, 64*(1), 35–46. doi: 10.1177/0022487112458801

*Shane, J. W., Binns, I. C., Meadows, L., Hermann, R. S., & Benus, M. J. (2016). Beyond evolution: Addressing broad interactions between science and religion in science teacher education. *Journal of Science Teacher Education, 27*(2), 165–181. doi: 10.1007/s10972-016-9449-4

*Shaw, J. M., Lyon, E. G., Stoddart, T., Mosqueda, E., & Menon, P. (2014). Improving science and literacy learning for English language learners: Evidence from a pre-service teacher preparation intervention. *Journal of Science Teacher Education, 25*(5), 621–643. doi: 10.1007/s10972-013-9376-6

*Shuls, J. V., & Trivitt, J. R. (2015). Teacher effectiveness: An analysis of licensure screens. *Educational Policy, 29*(4), 645–675. doi: 10.1177/0895904813510777

*Siegel, M. A. (2014). Developing preservice teachers' expertise in equitable assessment for English learners. *Journal of Science Teacher Education*, *25*(3), 289–308. doi: 10.1007/s10972-013-9365-9

Sleeter, C. (2014). Toward teacher education research that informs policy. *Educational Researcher, 43*(3), 146–153. doi: 10.3102/0013189X14528752

Soslau, E. (2015). Development of a post-lesson observation conferencing protocol: Situated in theory, research, and practice. *Teaching and Teacher Education*, *49*, 22–35. doi: 10.1016/j.tate.2015.02.012

*Steele, A. (2016). Troubling STEM: Making a case for an ethics/STEM partnership. *Journal of Science Teacher Education*, *27*(4), 357–371. doi: 10.1007/s10972-016-9463-6

*Stoehr, K. J. (2017). Building the wall brick by brick: One prospective teacher's experiences with mathematics anxiety. *Journal of Mathematics Teacher Education*, *20*(2), 119–139. doi: 10.1007/s10857-015-9322-y

*Strom, K. J. (2015). Teaching as assemblage: Negotiating learning and practice in the first year of teaching. *Journal of Teacher Education, 66*(4), 321–333. doi: 10.1177/0022487115589990

*Subramaniam, K. (2014). Prospective secondary mathematics teachers' pedagogical knowledge for teaching the estimation of length measurements. *Journal of Mathematics Teacher*

Education, 17(2), 177–198. doi: 10.1007/s10857-013-9255-2

Sugimoto, A. T., Carter, K., & Stoehr, K. J. (2017). Teaching “in their best interest”: Preservice teachers’ narratives regarding English learners. *Teaching and Teacher Education, 67*, 179–188. doi: 10.1016/j.tate.2017.06.010

*Sun, J., & van Es, E. A. (2015). An exploratory study of the influence that analyzing teaching has on preservice teachers' classroom practice. *Journal of Teacher Education, 66*(3), 201–214. doi: 10.1177/0022487115574103

*Swanson, L. H., & Coddington, L. R. (2016). Creating partnerships between teachers & undergraduates interested in secondary math & science education. *Teaching and Teacher Education, 59*, 285–294. doi: 10.1016/j.tate.2016.06.008

*Swars, S. L., Smith, S. Z., Smith, M. E., Carothers, J., & Myers, K. (2018). The preparation experiences of elementary mathematics specialists: Examining influences on beliefs, content knowledge, and teaching practices. *Journal of Mathematics Teacher Education, 21*(2), 123–145. doi: 10.1007/s10857-016-9354-y

*Talanquer, V., Bolger, M., & Tomanek, D. (2015). Exploring prospective teachers' assessment practices: Noticing and interpreting student understanding in the assessment of written work. *Journal of Research in Science Teaching, 52*(5), 585–609. doi: 10.1002/tea.21209

*Tallman, K. A., & Feldman, A. (2016). The use of journal clubs in science teacher education. *Journal of Science Teacher Education, 27*(3), 325–347. doi: 10.1007/s10972-016-9462-7

*Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Rowley, G., Peck, R., . . . Reckase, M. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. East Lansing, MI: Teacher Education International Study

Center, College of Education, Michigan State University.

*Thanheiser, E. (2015). Developing prospective teachers' conceptions with well-designed tasks: Explaining successes and analyzing conceptual difficulties. *Journal of Mathematics Teacher Education, 18*(2), 141–172. doi: 10.1007/s10857-014-9272-9

Thomas, M. A. M. (2018). ‘Policy embodiment’: Alternative certification and Teach For America teachers in traditional public schools. *Teaching and Teacher Education, 70*, 186–195. doi: 10.1016/j.tate.2017.11.011

*Thompson, J., Hagenah, S., Lohwasser, K., & Laxton, K. (2015). Problems without ceilings: How mentors and novices frame and work on problems-of-practice. *Journal of Teacher Education, 66*(4), 363–381. doi: 10.1177/0022487115592462

Thomson, M. M., & Palermo, C. (2014). Preservice teachers' understanding of their professional goals: Case studies from three different typologies. *Teaching and Teacher Education, 44*, 56–68. doi: 10.1016/j.tate.2014.08.002

*Trauth-Nare, A. (2015). Influence of an intensive, field-based life science course on preservice teachers' self-efficacy for environmental science teaching. *Journal of Science Teacher Education, 26*(5), 497–519. doi: 10.1007/s10972-015-9434-3

Tuck, E., & Gorlewski, J. (2016). Racist ordering, settler colonialism, and edTPA: A participatory policy analysis. *Educational Policy, 30*(1), 197–217. doi: 10.1177/0895904815616483

Tygret, J. A. (2017). The influence of student teachers on student achievement: A case study of teacher perspectives. *Teaching and Teacher Education, 66*, 117–126. doi: 10.1016/j.tate.2017.04.005

*Tyminski, A. M., Zambak, V. S., Drake, C., & Land, T. J. (2014). Using representations, decomposition, and approximations of practices to support prospective elementary mathematics teachers' practice of organizing discussions. *Journal of Mathematics Teacher Education, 17*(5), 463–487. doi: 10.1007/s10857-013-9261-4

*van den Kieboom, L. A., Magiera, M. T., & Moyer, J. C. (2014). Exploring the relationship between K–8 prospective teachers' algebraic thinking proficiency and the questions they pose during diagnostic algebraic thinking interviews. *Journal of Mathematics Teacher Education, 17*(5), 429–461. doi: 10.1007/s10857-013-9264-1

*van Ingen, S., & Ariew, S. (2015). Making the invisible visible: Preparing preservice teachers for first steps in linking research to practice. *Teaching and Teacher Education, 51*, 182–190. doi: 10.1016/j.tate.2015.07.001

*Wallace, C. S., & Brooks, L. (2015). Learning to teach elementary science in an experiential, informal context: Culture, learning, and identity. *Science Education, 99*(1), 174–198. doi: 10.1002/sce.21138

Weiland, I. S., Hudson, R. A., & Amador, J. M. (2014). Preservice formative assessment interviews: The development of competent questioning. *International Journal of Science and Mathematics Education, 12*(2), 329–352. doi: 10.1007/s10763-013-9402-3

*Whitacre, I. (2015). Strategy ranges: Describing change in prospective elementary teachers' approaches to mental computation of sums and differences. *Journal of Mathematics Teacher Education, 18*(4), 353–373. doi: 10.1007/s10857-014-9281-8

*Whitacre, I., & Nickerson, S. D. (2016). Investigating the improvement of prospective elementary teachers' number sense in reasoning about fraction magnitude. *Journal of Mathematics Teacher Education, 19*(1), 57–77. doi: 10.1007/s10857-014-9295-2

*Whitacre, I., & Nickerson, S. D. (2016). Prospective elementary teachers making sense of multidigit multiplication: Leveraging resources. *Journal for Research in Mathematics Education, 47*(3), 270–307. doi: 10.5951/jresematheduc.47.3.0270

Willegems, V., Els, C., Struyven, K., & Engels, N. (2017). Teachers and pre-service teachers as partners in collaborative teacher research: A systematic literature review. *Teaching and Teacher Education, 64*, 230–245. doi: 10.1016/j.tate.2017.02.014

*Wilson, R. E., Bradbury, L. U., & McGlasson, M. A. (2015). Integrating service-learning pedagogy for preservice elementary teachers' science identity development. *Journal of Science Teacher Education, 26*(3), 319–340. doi: 10.1007/s10972-015-9425-4

*Windschitl, M. A., & Stroupe, D. (2017). The three-story challenge: Implications of the Next Generation Science Standards for teacher preparation. *Journal of Teacher Education, 68*(3), 251–261. doi: 10.1177/0022487117696278

*Windschitl, M., & Barton, A. C. (2016). Rigor and equity by design: Locating a set of core teaching practices for the science education community. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 1099–1158). Washington, DC: American Educational Research Association.

*Wood, M. B., & Turner, E. E. (2015). Bringing the teacher into teacher preparation: Learning from mentor teachers in joint methods activities. *Journal of Mathematics Teacher Education*, *18*(1), 27–51. doi: 10.1007/s10857-014-9269-4

*Yeh, C., & Santagata, R. (2015). Preservice teachers' learning to generate evidence-based hypotheses about the impact of mathematics teaching on learning. *Journal of Teacher Education, 66*(1), 21–34. doi: 10.1177/0022487114549470

Zeichner, K., Bowman, M., Guillen, L., & Napolitan, K. (2016). Engaging and working in solidarity with local communities in preparing the teachers of their children. *Journal of*

Teacher Education, 67(4), 277–290. doi: 10.1177/0022487116660623

Zeichner, K., Payne, K. A., & Brayko, K. (2015). Democratizing teacher education. *Journal of Teacher Education*, *66*(2), 122–135. doi: 10.1177/0022487114560908

---

[1] We recognize the critical importance of early childhood experiences on young children's conceptions of STEM. However, early childhood teachers are trained in very different ways across states, and because of the prevalence of private preschools, many pre-K teachers are not certified. Pre-K teacher preparation is, therefore, complex and varied. In order to bound the synthesis, we have elected to leave out research on and measurement of pre-K teacher preparation. For readers interested in general research issues in pre-K, please see Horm, Hyson, and Winton (2013) as a starting place.

[2] We do not believe it is helpful to reinforce the historical distinction between quantitative or qualitative research because the distinction reifies the historical power dynamic between those two types of research, frequently carried out by different groups of researchers who may hold different positions in the research community. However, we need a way to communicate with the reader about studies that involve different numbers of participants, have samples that reflect the overall population differently, and have different goals. Thus, for simplicity of communication we use these categories.

[3] For some alternatively-certified STEM teachers who come to teaching as a second career after working in STEM fields, the workplace provides additional subject-matter expertise.