European Journal of Science and Mathematics Education, 2024, **12**(2), 175-199
ISSN: 2301-251X (Online)

OPEN ACCESS

Review Article

Qualitative research in science education: A literature review of current publications

Sabrina D. Stanley [1,2*]
0000-0001-7557-5640

William Boden Robertson [3]
0000-0003-0854-6641

[1] Department of Teaching, Learning, & Leadership, University of North Alabama, Florence, AL, USA
[2] Department of Curriculum & Instruction, The University of Alabama, Tuscaloosa, AL, USA
[3] Department of Educational Studies, The University of Alabama, Tuscaloosa, AL, USA
[*] Corresponding author: Sstanley5@una.edu

Citation: Stanley, S. D., & Robertson, W. B. (2024). Qualitative research in science education: A literature review of current publications. *European Journal of Science and Mathematics Education, 12*(1), 175-199. https://doi.org/10.30935/scimath/14293

ARTICLE INFO

Received: 27 Apr 2022
Accepted: 8 Feb 2024

ABSTRACT

This study analyzed articles from the last four years regarding how science education research is framed and discussed as qualitative research. The research question that guided this study was: *To what extent do qualitative secondary science teaching research publications reflect high-quality practices found in mainstream methodological texts?* The researchers utilized a systematic literature review methodology by (1) creating search terms based on the research question; (2) choosing relevant databases in which to search; (3) conducting the search and gathering articles; and (4) selecting articles based on inclusion criteria. The researchers chose "secondary education" and "science teaching" as search terms relevant to this study. Articles included in the review were peer-reviewed for credibility, available free online as full-text for accessibility, and available in English, which is the authors' first language. The researchers conducted three levels of screening on the full collection of articles–title, abstract, then methods, to efficiently narrow the large sample of qualitative science education research articles to a manageable and characteristic selection. The findings include that few articles addressed science teaching and learning with deep qualitative engagement. Some articles claimed to use specific qualitative methodologies without adequately expressing aspects of those methodologies, which lend support to the credibility, transferability, dependability, or confirmability of the articles such as the researchers' subjectivity or member-checking. Those studies that did are indeed diamonds in the ruff.

Keywords: science education research, qualitative research, systematic literature review, science education, science teachers

INTRODUCTION

Science education refers to the main science disciplines such as biology, chemistry, physics, and others. Science education research, however, is concerned with the ways science is taught by teachers, learned by students, and the affective factors that influence these phenomena. Duit (2007) defined three major domains of science education research:

(a) analysis of content structure,

(b) research on teaching and learning, and

(c) development and evaluation of instruction/instructional design.

How these domains are explored, described, or analyzed can vary according to the research questions and methods of the study. Another analysis on areas of study about science teachers in educational research include the domains of science teachers' knowledge, conceptions and beliefs, understandings of scientific

inquiry and the nature of science, pedagogical content knowledges, and knowledges of goals and curriculum (Avraamidou, 2014). Science teachers play a vital role in student learning of science, and research on science teachers is a part of that role. National Science Teacher Association (NSTA) maintains that continual inquiry into science teaching and learning promotes 21st century students' scientific literacy (NSTA, 2017). Similarly, the National Research Council (NRC) argued that teacher-based research is an important step in students' scientific proficiency and also leads to science education that is more equitable and inclusive (NRC, 2012).

Traditionally, educational research has been dominated by two approaches–qualitative and quantitative. The two approaches have underlying philosophical assumptions, which heavily influence how they ascribe to truth. While qualitative research typically rejects worldviews that ascribe to one truth (i.e., non-positivism), quantitative research typically seeks out a single truth (i.e., positivism). The difference in these assumptions stems from a contrast in the overall purpose of each approach and affects inquiries positioned under each paradigm differently. For example, while quantitative inquiries may be guided by "what" questions (e.g., what …), qualitative inquiries may be guided by "how" or "why" questions (e.g., how). Mixed methods research is the combined use of both quantitative and qualitative methods in a way to provide a better understanding of the research objectives than can be from a single method (Creswell & Plano Clark, 2011). .

Scientific research is a field steeped in quantitative traditions and research in science education has historically mirrored that paradigm. However, paradigms shift over time and articles in science education can today be quantitative, qualitative, or mixed methods. Researchers choose methods to carry out their qualitative research depending on the type, focus, and nature of the study. For example, when study participants were students, data collection focused on achievement, typically measured by quantitative standardized tests such as trends in international mathematics and science study. Studies related to student achievement tended to analyze quantitative data such as test scores (Libarkin & Kurdziel, 2002). In contrast, qualitative studies tended to focus on teacher beliefs, teacher practices, teacher development, teacher leadership or learner experiences in the classroom and collected data with interviews and observations (cf. Lundqvist & Sund, 2018; Vázquez-Bernal et al., 2021; Wen et al., 2021). Whereas quantitative approaches typically test theories by examining the connection or relationships among a set of variables, qualitative approaches tend to explore the meaning individuals ascribe to social phenomena (Creswell, 2014). As a result, quantitative approaches or quantitative data analyses may reveal meaningful interactions among variables and produce valuable findings from large data sets. By contrast, "qualitative research allows researchers to get at the inner experience of participants, to determine how meanings are formed through and in culture, and to discover rather than test variables" (Corbin & Strauss, 2008, p. 12).

Given the need for research on teachers (cf. NRC, 2012; NSTA, 2017), this study situates itself in the growing field of teacher-centered literature as a critical survey of the extent to which qualitative secondary science teaching research publications reflect high-quality practices as found in mainstream methodological texts. Despite an observable increasing frequency of publications that are qualitative in nature, there are no discernable guidelines from publications that define how qualitative research studies should be carried out. Moreover, differences across scholarly journals with respect to reviewers' training and professional opinions regarding the nature of qualitative research increases variability. The goal of this project is thus an attempt to understand more broadly the current state of the field's research by analyzing the qualitative research articles published in top-tier science education journals.

LITERATURE REVIEW

Previous systematic reviews of the literature informed the field about popular focuses within science education research. Chang et al. (2010), for instance, systematically analyzed the trends in science education research. Through a content analysis method of segmenting, clustering, and visualization, the authors analyzed 1,401 articles from four top tier journals for the years 1990-2007 and identified nine topics in science education research. However, by choosing to analyze a large number of articles, the author's methodology could not possibly accommodate a detailed analysis of each article's methodology. Furthermore, Chang et al.'s (2010) stated purpose of analysis was to identify topics, trends, and contributors to the field of science education research so that novice researchers may more easily understand which areas in the field need to be further explored, which excluded a motivation to analyze each article individually.

Karampelas (2021) similarly sought to better understand trends in education research and reviewed a large number of articles (n=6,504) published over the ten-year period of 2010-2020. Whereas Chang et al. (2010) focused on journals that emphasized science education, Karampelas' (2021) analysis looked at a range of topics in educational research articles published in four journals for evidence of attention given to science education research. They found that of the thousands of articles they screened, 400 were based on the topic of science education. However, their analysis did not extend to identifying trends *within* science education, rather it reported on trends within education research as a whole.

Lin et al.'s (2018) examination of research trends in science education from 2013-2017 was similar to both Chang et al. (2010) and Karampelas (2021) in providing the field with an analysis of article trends through systematic analysis. The author's study was the latest in a series of studies carried out by a group of scholars (see Lee et al., 2009; Lin et al., 2013; Tsai & Wen, 2011) on the same three journals (e.g., *Science Education, Journal of Research in Science Teaching, and International Journal of Science Education*) every four-five years. Of the four papers published by this group, two reported a breakdown of paradigmatic alignment and Lin et al. (2013) of articles analyzed. Lee et al. (2009) reported that 10 out of 15 empirical studies published from 1998-2002 were qualitative. The group's later publication, Lin et al. (2013) reported that of articles published from 2008-2012, three out of seven studies were qualitative. However, none of this group's articles have analyzed how each paper carried out its methodology.

Two published works titled *The Handbook of Research on Science Education* contained comprehensive syntheses of science education research. Volume one categorized its articles under the following headings:

(a) science learning,

(b) culture, gender, society, and science learning,

(c) science teaching,

(d) curriculum and assessment in science, and

(e) science teacher education (Abell & Lederman, 2007).

Volume two expanded the volume one list to also include theory and methods of science education research as well as diversity and equity in science learning (Lederman & Abell, 2014). The articles included in these two volumes of the science education research handbooks indicated the types of research publications that were widely accepted as the best in the field. The handbook categorizations were similar to those found in other literature reviews (c.f. Chang et al., 2010, Karampelas, 2021; Lee et al., 2009; Lin et al., 2013, 2018; Tsai & Wen, 2011), however, the handbooks did not include an analysis of paradigmatic orientation or methodology.

While these literature reviews focused on qualitative secondary science studies more broadly, their findings underscored a need for closer analysis. For example, Chang et al.'s (2010) survey of publications from 1995-2005, the authors found that "researchers were becoming small-scaled in research design while more qualitative data collection methods were used" (p. 317). This is a similar finding to that of Karampelas (2021), who wrote that qualitative methodologies have become preferred due to a growing concentration of researchers on classroom teaching practices, which "researchers prefer to carry out [with] empirical research that takes place in a classroom in the form of an action research case" (p. 10). Because of this increasing number of researchers deploying qualitative methodologies, there is a need to analyze the individual methods and methodological orientations of authors engaging more closely in this work. Moreover, it was in this analytic deficiency that this article positioned itself so that the field may know not only what broad trends were present in research, but more specifically what methods were being used and how they were being deployed.

Research Question & Purpose

The purpose of our systematic review is to go beyond previously completed reviews, which have reported paradigmatic orientations and more fully explore which methods are used and *how* they were used. This systematic literature review analyzed articles over the last four years that utilized qualitative methodologies that centered science teachers, answering the call by the NRC (2012) and NSTA (2017). The research question

that guided the study through systematic literature review was: *To what extent do qualitative secondary science teaching research publications reflect high-quality practices found in mainstream methodological texts?*

CONCEPTUAL FRAMEWORK

To answer this research question, we sought to develop a conceptual framework around how qualitative methodologies have been deployed in secondary science teaching research. We developed this conceptual framework around what we consider to be most applicable to secondary science studies that focus on the teacher. Moreover, while previous studies have focused on qualitative methodologies broadly, we developed our conceptual framework to engage with a smaller number of studies more closely. Toward this end, this conceptual framework functions to clarify our intent and methodological choices.

Qualitative Studies

Qualitative methodology is characterized as being descriptive and dealing with phenomena in a deeper or more individualized way (Merriam, 2009). In qualitative studies, researchers are the designers of the study, the collectors and analyzers of the data, and in some cases the facilitators in the programs being studied. Maxwell (2013) advocated that, no matter the level of involvement, the researcher is an instrument in the qualitative study, representing a strong departure from quantitative studies that consider researchers to be neutral. Additionally, quantitative studies collect different forms of data than qualitative studies as they "seek to recreate the contextual setting as a framework that can be analyzed and understood. By necessity, qualitative research often consists of as much data as possible, including detailed field notes, tape and video transcripts, and written documents" (Libarkin & Kurdziel, 2002, p. 80). Toward this end, qualitative studies tended to utilize some form of interview and observation. Qualitative researchers often coded this data looking for patterns in order to draw conclusions and imbue themselves in the research (Saldaña, 2016).

Qualitative researchers thus have differing practices from those found in quantitative studies. For example, in a quantitative study, reliability and validity have long been used as standards of consistency or accuracy of the measures utilized in the study. A qualitative study, however, is evaluated on the basis of credibility, transferability, dependability, and confirmability (Lincoln & Guba, 1985). However, for a variety of reasons (e.g., journal reporting standards, researchers' professional writing styles, etc.), these individual categories may not be explicitly addressed leading to confusion or misinterpretation by readers.

Credibility

The logic behind researcher choices must make sense to the reader and lend a sense of credibility to the researcher; "the basic notion with credibility is that both the readers and participants must be able to look at the research design and have it make sense to them" (Jensen, 2008b, p. 139). Had we chosen to focus on articles in the area of K-12 history education, this would represent a disconnect in the logic of our choice causing a credibility crisis when readers attempt to understand our research design. Jensen (2008b) recommended five ways that researchers can improve their credibility: spending enough time in the field with participants; angles: utilizing various perspectives to analyze data to gain a more holistic view; colleagues: reach out or partner with other researchers in the field to review your analysis and findings; triangulation: utilize multiple sources as well as multiple methods to gather data; member checks: enlist help from participants to ensure your analyses are accurate. Other scholars have argued of increased credibility through prolonged engagement (Korstjens & Mosler, 2018), the use of member checking to confirm participant responses (Anderson, 2017), and triangulation through the use of multiple data points (Stake, 1995; Steinke, 2004).

Transferability

Researchers generally choose study participants with the "inherent notion that they somehow represent the entire population" (Jensen, 2008d, p. 886). This is so that readers may make connections from a researcher's study to other applicable contexts. However, this is not true of all types of qualitative research. For example, Stake (1995) argued that while it may be beneficial to select typical cases for case study research,

atypical cases may help us understand matters that are often overlooked. Jensen (2008d) wrote that qualitative researchers could increase the transferability of their research by

(a) ensuring the study's context and participants are closely intertwined through purposeful sampling,

(b) conveying the study's context using thick description so that readers are given a full account of all aspects of the research, and

(c) ensuring that each research question is answered appropriately.

Dependability

One unavoidable reality of research in general is that the context of research may change over time. Values for quality are contextual (Tracy, 2010). This is important for qualitative research as it may produce findings that are highly sensitive to environmental contexts. This may also happen as unexpected forces arise at some point during the research process. Jensen (2008c) wrote that dependability "recognizes that the research is evolving and that it cannot be completely understood a priori as a singular moment in time" (p. 209). To increase a study's dependability, researchers should give readers enough information about the structure of the research project–including methodological choices–so that others may attempt to replicate it. With this information, other researchers can account for changes in other contexts as they attempt to replicate the study in that context.

Confirmability

One of the implicit ideas of a published research article is that its findings should be confirmable by other researchers. Readers must be provided enough evidence to support a researcher's conclusions, findings, or implications. Jenson (2008b) defined confirmability as, "the degree to which the results of the study are based on the research purpose" (p. 113). Anderson (2017) proposed a "communication of methodological awareness evidenced by an audit trail as a standard of quality in qualitative research" (p. 127). An example of this would be to include coding examples in the write up, asking participants to review coding processes and analyses, and the use of multiple data points through triangulation (Korstjens & Mosler, 2018).

METHODOLOGY

Systematic Literature Review

Cooper (2017) characterized a systematic review as containing four steps:

(a) create search terms based on the research question,

(b) choose databases with which to conduct the search,

(c) conduct the search and collect articles, and

(d) select articles based on inclusion criteria.

Following this guidance, we first established the search terms "science teaching" and "secondary education." Next, we chose databases based on what was available through our institution. The databases chosen were "ERIC EBSCO Host," "APA PsycInfo," and "Education Full Text (H. W. Wilson)." We then used the search terms in the databases. Furthermore, we selected criteria designed to limit the search to peer-reviewed journal articles published since 2018, available online as full text, and published in English. These criteria were selected for several reasons. We believed that articles published since 2018 would ensure the literature returned was current and relevant. Due to the scope of this project, full text versions of each article were necessary for in-depth analysis. Additionally, articles available to the researchers for no charge were selected due to funding. English language articles were chosen because English is the first language of the research team. This search yielded 928 non-duplicated articles. For the sake of comparison, we searched ERIC EBSCO Host with the same terms without the additional criteria. This resulted in 1,139,711 individual articles, an unmanageable quantity for two researchers..

We repeated the article search with the Web of Science database, which accesses open-source journals as an effort to broaden the selection of articles. We set similar search criteria and limitations–open access, peer-reviewed journal articles from 2018-2022 published in English. In this database search we were able to select

the following categories–education educational research and education scientific disciplines. The Web of Science search provided 752 additional articles to screen.

We next exported all 1,680 articles included from both searches (i.e., 928 from search 1 and 752 from search 2) to ProQuest RefWorks to gather pertinent information such as the author(s), title, citation, and abstract. Next, we transferred the article information to an Excel Spreadsheet for sorting and screening. We used the following criteria for inclusion: The study was

(a) primarily focused on science education,

(b) in a secondary education setting,

(c) qualitative studies, and

(d) pertained to classroom teachers and teaching practices.

We took the 1,680 articles from the search results through three phases of screening.

First, we screened the title of each article according to the inclusion criteria listed above. Articles were eliminated that focused on non-education subjects such as cancer, diet, mental health, or sexuality as they were beyond the scope of the inclusion criteria to primarily focus on science education. Some articles were eliminated at this phase of screening because they addressed non-science education topics such as mathematics, English language learning, physical education, or literacy. Articles identifying both mathematics and science were not eliminated during the title screening. Topics identified as STEM but not science specifically were not eliminated during this phase. However, some STEM-related articles were discarded such as those addressing only engineering, robotics, or computer science. Additionally, all articles with elementary or collegiate settings that were extraneously included through the database searches were eliminated at this stage as they did not fit within the criteria of secondary education settings. A total of 1092 articles were eliminated by title screening, leaving 588 articles to continue into the abstract screening. Subsequently, we conducted an abstract screening employing the same inclusion parameters–qualitative studies of secondary science education teachers. Eliminated articles based on the abstract included any non-qualitative studies such as quantitative or mixed methods research studies, studies developing or evaluating measurement tools or curricula, which do not limit their focus to the science teacher, meta-analyses, and literature reviews. The abstract screening eliminated an additional 392 articles. Finally, we scanned the methods section of 196 remaining full text articles to identify qualitative, empirical studies, which were based on observed, actual experiences as opposed to theory. The methods section screenings further illuminated articles to exclude any that were mixed methods studies that did not originally present as such in terms of Creswell's (2014) definition for mixed methods research. Also excluded were more program or curriculum design studies that slipped through notice from the abstract screening. Additionally, studies involving preservice teachers were eliminated in favor of limiting the search to in-service teachers to align with the research question. At this point, we deselected qualitative studies based solely on questionnaires as those studies were not likely to demonstrate rich qualitative engagement such as interpersonal interactions through observations and interviews. Additionally, articles lacking open or institutional access were eliminated to ensure ready availability for the project. The methods section screening resulted in elimination of an additional 171 articles. This detailed three-level screening process left us with a total of 25 eligible articles for the literature analysis to address the research question. PRISMA diagram (Moher et al., 2009) illustrates our search and screening process (**Figure 1**).

Following screening, we read each of the 25 included articles (see **Appendix A**) and coded them manually. We approached coding in a deductive manner as we had a predetermined set of variables related to qualitative research. These variables were determined through meetings in which we discussed what we felt were indicators of rich quality in qualitative research. We also consulted the literature as well as colleagues in the field to verify and more fully understand these indicators. After establishing the indicators as variables, we created a matrix to track article features. The matrix variables were, as follows:

(a) role of researcher,

(b) methodology,

(c) data collection,

(d) time durations,

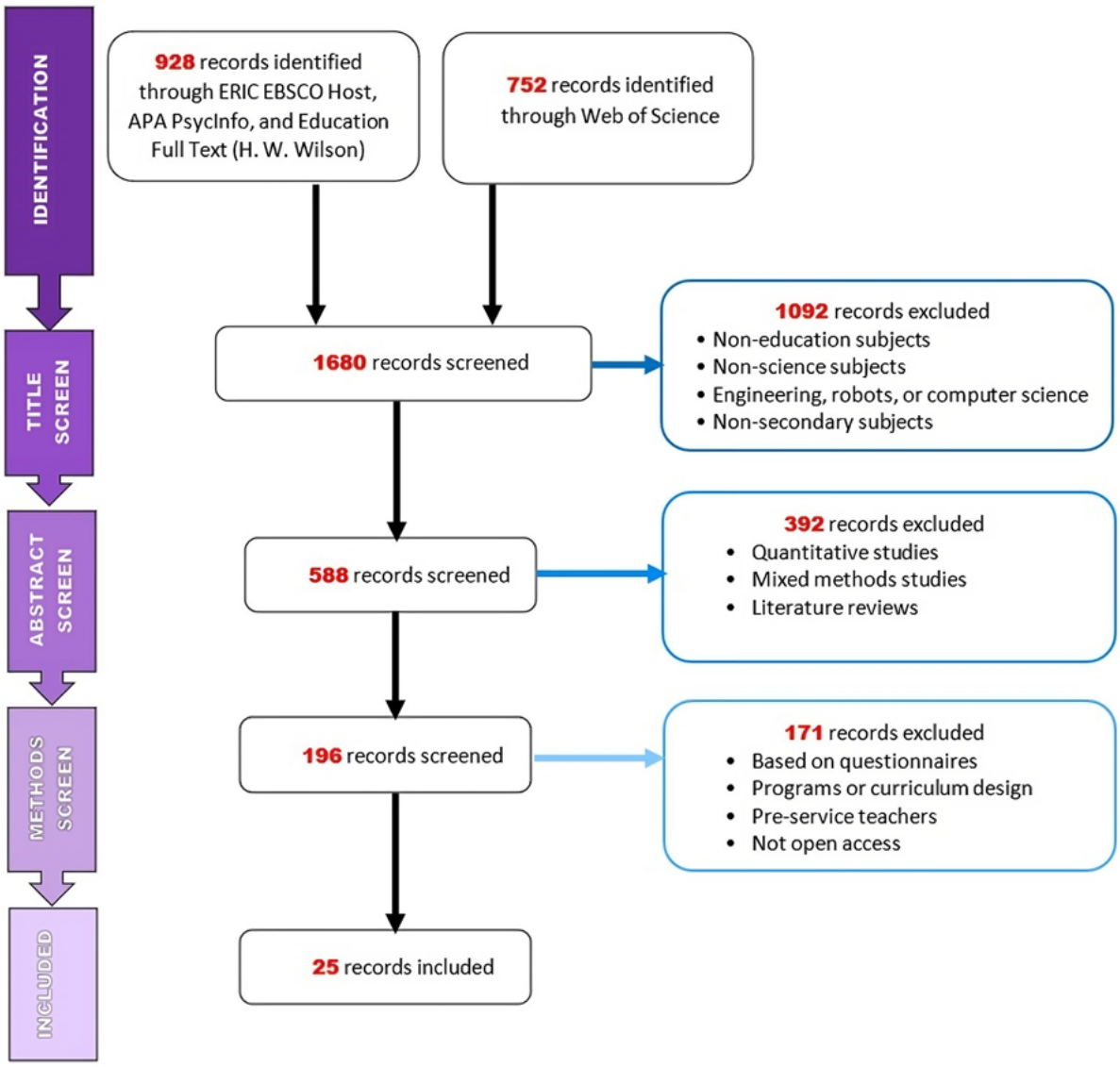

Figure 1. PRISMA diagram (Adopted from Moher et al., 2009)

(e) triangulation,

(f) member-checking, and

(g) coding method (see **Appendix B** for the coding explanation).

These variables were chosen based on our understanding of research in science education as well as how we understand qualitative research with respect to credibility, transferability, dependability, and confirmability. For example, because this project sought to understand the methodological approach of articles that center the experiences of science teachers, the first variable coded for was whether the researcher was a participant or observer helps the reader understand the perspective of the researcher as well as how their perspective impacts the study. We also recorded the methodology explicitly stated by the author as it allowed us to understand the researcher's approach within the broader context of mainstream methodological texts as well as how they stated and approached data collection. The variables of time spent in the field and triangulation through multiple data sources not only give credibility to the researcher's data collection, but also the confirmability of their analyses. Similarly, member variables were recorded as enlisting the help of research participants to confirm findings also increases credibility. We recorded information related to coding as a researcher's transparency in revealing their coding method helps to increase their article's transferability as it ensures the reader that each question was appropriately answered. Similarly, this also increases an article's dependability as other researchers may attempt to replicate an author's inquiry. A full list of the variables and descriptions are found in **Appendix C**.

Credibility, Transferability, Dependability, & Confirmability of Current Study

Credibility

In order to increase the credibility of our study, we adhered to writings from Jensen (2008b) and relied on our collective understanding of credibility developed from qualitative methods coursework we had completed during our doctoral studies. The methodology used for this systematic literature review was presented at the Southeastern Universities Graduate Research Symposium (SUGRS) Conference, an annual regional conference hosted by The University of Alabama's College of Education. We received rich detailed feedback about the methodology from reviewers who are current science education faculty that regularly engage in scholarly research and publication. This feedback was considered as we moved forward with our study. For example, instead of reporting variables like triangulation as dichotomous (i.e., yes or no), we recorded how a researcher used triangulation (i.e., using focus group data, interview data, and journal entry data). Additionally, the articles chosen for this systematic review were rigorously analyzed: first independently by each author, then as a group. When we met to discuss the articles we talked through disagreements with each author's analysis, and each article was checked several times to ensure consistency in our analysis.

Transferability

Our systematic literature yielded findings from journals with high impact factors and broadly representative of science education scholarship (Jensen, 2008d). Additionally, we were purposeful in sharing as many rich methodological descriptions as possible within the limits of the journal's wordcount.

Dependability

We worked to increase our study's dependability by explicitly describing all steps and search factors in order to provide readers the information necessary to replicate this research project. The values that impact beliefs in qualitative research have changed dramatically over time and these values will surely change in the future. We believe that by describing our process and values in as much detail as possible as well as connecting them in the literature will help future researchers understand our findings so that they may replicate it later when reporting standards inevitably change.

Confirmability

To convey a sense of confirmability to our study, we strived to be as thorough as possible when describing our methods. Some methodological decisions we made in this study to support its confirmability include descriptions of the coding themes, providing a PRISMA diagram of the literature search process, and including research support of qualitative methodologies used.

FINDINGS

As a review of qualitative science education research, our findings were guided by our coding process and are presented as findings based on the coded variables:

(a) role of researcher,

(b) methodology,

(c) data collection,

(d) time durations,

(e) triangulation,

(f) member-checking, and

(g) coding method.

A full coding matrix is also provided (**Appendix C**). We do not propose that any one feature of the coding has precedence over another.

Variables

Role of researcher

Researchers positioned as outside observers constituted the bulk of the articles (n=21). However, four of the articles featured the researcher as a participant (Hordvik et al., 2021; Keiler, 2018; King & Pringle, 2019; Velasco et al., 2021). Only two of the publications that we reviewed described the impact of the researcher in the study (e.g., Gardner & Tillotson, 2019; Hordvik et al., 2021). When authors considered researcher influence, it was framed in terms of their efforts in coding the data rather than their interactions with the participants (e.g., Brown & Bogiages, 2019; Fitzgerald et al., 2019).

Methodology

We found that authors identified their study's methodology beyond stating they were qualitative in all but four articles. The methodologies explicitly stated by the authors were case studies, narrative or discourse analysis, phenomenology, and one observational study.

Case study: Case studies represented the majority of the authors' chosen methodologies. Some authors described their case studies as descriptive (e.g., Gardner & Tillotson, 2019; Wilson, 2021), longitudinal (e.g., Dogan et al., 2020; Vázquez-Bernal et al., 2021), or multi-case (e.g., Kirmaci et al., 2019; Vossen et al., 2020). Brown and Bogiages (2019) called their study an instrumental case study while Velasco et al. (2021) identified their methodology as an embedded single-case study. Two studies did not identify any particular methodology but could have been framed as case studies based on our understanding of what case studies are and how they were identified by other authors in the literature review (i.e., Litman & Greenleaf, 2018; Vale et al., 2020). On the other hand, several of the studies that explicitly stated case study as their methodology, did not mention key components of a case study. For example, only eight of the 13 self-identified case studies mentioned triangulation, a common hallmark of case studies found in methodological literature in education research (Merriam, 2009; Stake, 1995).

Narrative/discourse analysis: Two articles stated they used discourse analysis methodology. Andrée and Hansson (2021) sought to understand teacher agency by analyzing talk among teachers from five focus groups as they evaluated curriculum materials provided by commercial agencies. The researchers stated that discourses are socially embedded and appropriate for analyzing how teachers talk about these resources. King and Pringle (2019) acknowledged the world was culturally and socially defined, therefore the students' narratives were important for giving their own account. Each researcher created narratives from participants' interview transcripts while participants wrote narrative accounts of their own experiences in STEM education.

Phenomenology: Three authors identified their studies as phenomenological. Birth et al. (2018) used a phenomenological methodology to understand perception of physics teachers of professional development. Strachan (2020) similarly used phenomenology to explore the experiences of two African American science teachers "to determine the connection between what is being perceived and how it is being experienced" (p. 228).

Data collection

In the articles analyzed, data collection largely took the form of interviews, focus groups, and observations. Articles differed in the way they carried out a specific collection method. For example, Berge et al. (2020) analyzed body language in video recordings, while Gardner and Tillotson (2019) made classroom observations of teachers. However, we classified both of these data collection methods as *observations*.

Interviews: 19 articles (76%) identified semi-structured interviews as a data collection method. Two of these articles stated they used a distinct approach to interviewing, using what they called video-stimulated interviews (i.e., Overman et al., 2019; Vale et al., 2020). In these articles, participants and interviewers watched video recordings together of the participant teaching lessons to focus the discussion on critical moments and to prompt further reflection by the participants. Studies that engaged in interviewing, discussed the value of interview as a data collection method in various ways. For example, Gardner and Tillotson (2019) stated that "teacher interviews provided historical perspectives of the [STEM] model" of the institution (p. 1288). Another perspective on interview was from Nixon et al. (2019) who acknowledged that interviews have limits as "these

interviews only approximated tasks of teaching, as they were removed from the context of teaching. In the complexity of the actual teaching practice, it is more challenging to identify the knowledge teachers are using" (p. 155).

Authors used interview transcripts in a variety of ways. Several studies briefly stated that interviews were recorded and transcribed verbatim (Andrée & Hansson, 2021; Birth et al., 2018; Hordvik et al., 2021; Vossen et al., 2020; Wen et al., 2021). Other researchers attended to the transcriptions in more detail in their publication. An example, Berge et al. (2020):

> All presentations were first transcribed verbatim and read to obtain an overview of the data … In order to be able to compare and contrast expected and unexpected patterns in teacher-student interaction the three presentations, all thematically varied and comprehensive, were transcribed a second time to capture important body language such as what the teacher drew on the whiteboard in the classroom (p. 68).

Likewise, King and Pringle (2019) described their transcription process this way:

> In crafting the first draft of the profile, we transcribed the first interview word-for-word with the coughs, sneezes, giggles, pauses, and idiosyncrasies. We kept the original transcript but started a new document, where we deleted all of the interview questions from the transcript so that only the girls' words remained in the document (p. 553).

Although all the articles included quotes from the interview transcripts, each author described the process differently. Studies reported a variety of means of conducting interviews, transcribing interviews, and the length of interviews. As examples:

(a) Lundqvist and Sund (2018) documented "the main data were collected by means of three group interviews, each lasting approximately 75 min" (p. 359),
(b) Velasco et al. (2021) described "the individual interviews were 45 min, conducted and recorded using the video conference application Zoom" (p. 441), and
(c) Walan (2020) simply put "the interviews were semi-structured, audio-recorded and transcribed" (p. 434).

Focus groups: Focus groups were a data collection method described in eight of the qualitative studies. Each author described their rationale for using focus groups differently. For example, Keiler (2018) used focus groups "to prioritize teachers' perspectives about their experiences" (p. 5). Through a different lens, Andrée and Hansson's (2021) "study is based on an analysis of discursive practices employed by teachers in focus group conversations" (p. 357).

Observations: Observations were used as a data collection method in 11 of the 25 publications (e.g., Berge et al., 2020; Dogan et al., 2020). Their descriptions and bases for using observations varied. Wilson's (2021) protocol included how the observation protocols were developed including their setting, social climate, program and unplanned activities, and non-verbal communication. Other articles were less explicit when describing observation protocols. For example, Litman and Greenleaf (2018) explained their observation protocol stating, "classroom observations were protocol driven. The observation and analytic protocol focused on three aspects of the lesson: texts, classroom activities, and classroom culture" (p. 111). Similarly, Berge et al.'s (2020) article stated they "document[ed] classroom activities through video recordings" (p. 67).

Time durations

20 (80%) of the articles explicitly cited times from their research process; these reported the time spent observing or the length of time of the interviews. For example,

(a) Dogan et al. (2020) reported interviewing their participant for three years "in the fall of each year for periods of about 1 to 1.5 hours" (p. 87);
(b) Gardner and Tillotson (2019) recorded a total of 1383 min of classroom instruction out of four classrooms multiples times per week over three months; and

(c) Keiler (2018) held interviews and focus groups that lasted from 20 to 90 minutes four times a year for three years of the program under study.

Triangulation

Only half of the case studies indicated they utilized triangulation in their study design and analysis. Without explicitly stating so, Kirmaci et al. (2019) thoroughly described appropriate data triangulation:

> We constantly compared and cross-checked the data that were collected through individual and focus-group interviews and participant observations to develop better understanding of and provide multiple sources of evidence of how teachers' participation in the program influenced their perspectives and practices (p. 15).

Other examples include King and Pringle (2019) who stated they used triangulation in their narrative inquiry study for "robust counter-stories" (p. 539), and Navy et al. (2020) employed triangulation in their study by using multiple data sources collected by different researchers to "ensure validity and reliability of the findings" (p. 191).

Member-checking

Only 11 articles (44%) reported member-checking with the participants. King and Pringle (2019) used member-checking "to ensure that the narrative being written accurately depicted the girls' perceptions and experiences … The Black girls, as participants in the research, were elevated to the position of co-researchers and knowledge generators through the co-construction of their counter-stories" (p. 553-554). Their study contained multiple phases in which participant input influenced the direction of the study. Another example was Vázquez-Bernal et al. (2021) who involved the participant and wrote, "finally, in phase 3 (2011-2019), Marina was given the opportunity to read and write as narratives a major part of the reports elaborated by the researchers in the first two phases" (p. 5).

Coding

We analyzed each article's data analysis section based on our understanding of deductive and inductive coding and further informed by Saldaña (2016). We have included these results in the coding chart in **Appendix C**. Each article reviewed utilized coding as a method of data analysis. We found these articles were implicitly or explicitly coded either deductively (n=11), inductively (n=11), or both (n=3). However, authors differed qualitatively in how their coding was reported. For example, Navy et al. (2020) stated, "coding was partway between a priori and inductive approaches" (p. 191). Additionally, the authors explicitly state that the deductive coding in their first stage was based on "three primary types of resources (human, material, social)," which reflected an integration of the article's conceptual framework into their coding (p. 191). Brown and Bogiages's (2019) codes were based on a previous pilot study and the authors included a rich (~450 words) description of their coding process. Walan's (2020) coding came from a previous study by Pringle et al. (2015) that "used selective parts of TPACK [technological pedagogical content knowledge] to code data" (p. 432). The author stated, "I also decided to use only certain parts of TPACK" (p. 432). However, these codes were not revealed explicitly. Hordvik et al.'s (2021) article did not utilize coding in the traditional sense as the author's theoretical framework rejects static conceptualizations of codes for analysis. Despite this, the author's description of their analysis was rich and highly detailed. Some authors identified computer software, such as NVivo or Atlas.Ti, for organizing their codes in logical ways (c.f. Navy et al., 2020; Velasco et al., 2021; & Wilson, 2021). Fitzgerald et al. (2019) used a digital application to produce graphical visualizations but manually coded their data.

DISCUSSION

This systematic literature review sought to understand how qualitative secondary science teaching research publications reflect high-quality practices found in mainstream methodological texts. We found that the studies examined in this systematic literature review differed not only in their alignment to qualitative

methodologies, but also in the way data collection, time durations, triangulation, member-checking, and coding were reported. This analysis led us to make the following arguments.

Secondary Science Teaching Qualitative Methodologies

With respect to methodology, many studies reported themselves as case studies. Yazan (2015) wrote that case study was the most widely used qualitative research methodology. This research design is characterized by a small sample size delineated by a boundary and triangulation of data through multiple data sources. Stake (1995) described case studies as purposefully highly contextualized. They wrote "the real business of case study is particularization, not generalization. We take a particular case and come to know it well, not primarily as to how it is different from others but what it is, what it does. There is an emphasis on uniqueness…on understanding the case itself" (Stake, 1995, p. 8). However, some of these studies do not follow the characteristics of case study laid out by prominent case study methodologists (i.e., Stake, 1995; Merriam, 1998, 2009) and more closely resemble a standard approach to qualitative inquiry. Merriam (1998) asserted, "the term case study is not used precisely; it has become a catchall category for studies that are clearly not experimental, survey, or historical. And to a large extent, the term has been used interchangeably with other qualitative research terms" (p. 43). Stake (1995) went into great detail about what can be considered a case and what should not. More specifically, he stated that "people and programs clearly are prospective cases. Events and processes fit the definition less well, and studies of them are less likely to capitalize on the methods [of case study]" (p. 2). Stake's argument was aligned with other scholarly arguments around case study's bounded system (i.e., Merriam, 1998; Smith, 1979). Bounded systems serve a methodological function, focusing the researcher's attention to the case under study. In this way, the case study boundary functions as a set of metaphorical blinders, giving the researcher laser focus on what is under study (Robertson & Yazan, 2022). Simply studying one particular entity does not necessarily make that study a case study. For example, Brown and Bogiages (2019) presented a case study without the earmark features of a case study such as triangulation or defined boundaries. Of the 25 articles within this literature review, 13 identified as case studies. However, only Keiler's (2018) article contained elements of careful case study engagement as they reported time spent in field, triangulation, and member-checking. Without the components of a case study (i.e., clearly defined case, case boundary, triangulation, and member-checking), what many report as "case studies" are more accurately general approaches to qualitative inquiry.

Phenomenological research designs are broadly applicable as they are a more general way to qualitatively explore the views of the participants. Phenomenology is the study of phenomena, and the personal experiences of people. Vagle (2014) explained the purpose of phenomenology is "to study what it is like as we *find-ourselves-being-in-relation-with* others … and other things …" (p. 20). As opposed to case study, a phenomenological study is not accompanied by the same elements such as boundaries. Instead, a phenomenological study can be a free-form, in-depth look at personal experiences. Groenewald (2004) wrote that phenomenology was less prescriptive since imposing a strict method would compromise the integrity of the phenomenological methodology. Birth et al. (2018) used the phenomenological approach to their study "because of its emphasis on the phenomenon of the study" (p. 91), that being the teachers' perceptions of a particular program. Further, they justified that such methodology uses an emerging qualitative approach involving data collection in the natural setting and data analysis of patterns and themes (Birth et al., 2018). Gardner and Tillotson (2019) situated their study as both a phenomenology and a case study claiming the "phenomenological investigation emerged as an appropriate research method" (p. 1287). Because phenomenology is less restrictive with respect to the elements that define what it is, it proves to be useful for secondary science researchers who seek to understand the views and experiences of participants.

Reporting Standards of Qualitative Secondary Science Research

This systematic literature review yielded articles demonstrating a range of engagements with qualitative research methods. Hordvik et al. (2021) and Birth et al. (2018) presented very qualitatively oriented articles. For example, Hordvik et al. (2021) described the data production in this way: "The multiple layers of qualitative data provided varied sources of experience and modes of expression, thus helping enhance the trustworthiness of the findings" (p. 5). Likewise, Keiler (2018) discussed qualitative concerns such as

minimizing bias, including participants with diverse views and experiences, and triangulation of data. These articles engaged deeply with the overarching goals of qualitative research.

Data collection through interviewing is ubiquitous in qualitative research studies. Despite its widespread usage, variations in transcription practices may impact data analysis and reporting of findings. Researchers may manually transcribe interviews or utilize a variety transcription services–human or artificial intelligence–assisted. Additionally, transcription practices may only reproduce spoken words, whereas others may reproduce utterances like "um's," "ah's," throat clearings, long pauses, and a host of other audible sounds. These utterances, which may lack meaning on the surface, can often be used to determine emotional context or aid in conversation analysis (Poland, 2011). Despite these varying practices, some scholars avoid describing how their transcription was produced. For example, Berge et al. (2020) stated, "all presentations were first transcribed verbatim and read to obtain an overview of the data" (p. 68). Additionally, when the authors included participants' utterances, they were reproduced with narrative conventions of syntax and grammar, which differ greatly from spoken speech. The word "verbatim" means word for word in both American English and British English varieties (Verbatim, n. d., 2023). Contrasted with Berge et al. (2020), Andrée and Hansson's (2021) transcripts are filled with "like's" and ellipses to represent long pauses, more closely resembling spoken utterances.

Reporting research durations is not standardized; however, qualitative research studies should report how long interviews last, the length of time for an observation, or the time of a focus group session. What is often referred to as "time spent in field" improves both transparency and credibility (Jenson, 2008a). To be clear, methodological texts do not prescribe a perfect amount of time, but researchers should make the amount of time spent in the field known to the reader. For example, Dogan et al. (2020) indicated they wanted to uncover the epistemological underpinnings of a teacher's beliefs. The author utilized various observation protocols over a three-year span and provided the length of time for each annual interview. Given the context of what the author sought to uncover, it makes sense that they undertook a prolonged engagement in the field. Had the author sought to investigate the same phenomenon but only spent one month in the field, the reader's sense of the author's credibility would be undermined.

Triangulation can be a powerful tool in qualitative research. Some researchers may wish for findings to converge to amplify their confirmability, while others, particularly in case study research, may desire a richer analysis that comes from weighing multiple data sources. According to Flick (2022), Merriam (1988), and Stake (1995), triangulation is an integral part of case study research as it allows a researcher the opportunity to collect data using multiple methodologies, which have differing strengths and weaknesses so that when combined these deficiencies may be overcome. However, despite its importance, scholars differ in their use and purpose of triangulation in their data analysis. Velasco et al. (2021) explained their use of triangulation in their case study was to "to check the consistency of findings by using focus group and document data as secondary sources to inform analysis of the individual interviews, which were the primary data source" (p. 443-444). Kirmaci et al. (2019) similarly stated that they used triangulation to enhance trustworthiness. However, some studies that claimed to be case studies did not use triangulation at all (e.g., Brown & Bogiages, 2019; Wen et al., 2021).

Member-checking is a process related to research credibility. King and Pringle (2019) utilized extensive member-checking with their young participants as they "diligently listened to the girls to co-construct their counter-stories with authenticity" (p. 550). The authors sought to increase Black girls' engagement with STEM and utilized a methodology informed by critical race theory. Meaningfully including their participant into the project at multiple stages and empowering them as co-researchers not only increased their study's confirmability, but it also aligned with the goals of the authors' theoretical framework. Vázquez-Bernal et al. (2021) provided their participant the "opportunity to read and write as narratives a major part of the reports elaborated by the researchers in the first two phases" (p. 5) over eight years of participation. In this way, the researchers continuously performed member-checking, increasing the credibility of their findings. In contrast, Dolfing et al. (2020) presented their research as a case study yet made declarations about their data analysis that do not align with methodological texts. The authors stated,

> The interpretations of the results of both authors, in general, were comparable. Member checks were not performed during the program, as they would have had too great an influence on the

> process of sense-making, whereas, after the program, member checks would result in teachers' rational responses concerning their intuitive, emotional process of sense-making (p. 145).

As Dolfing et al.'s (2020) stated purpose of their research was to understand teachers' process of sense making, the researchers' analysis could have been richer from including participants' thoughts of the researchers' analysis. Stake's (1995) rationale for member checks was, "I often do not have all my facts straight and I need help … I think I can say that all my reports have been improved by member checking" (p. 116).

Each article reviewed utilized coding during data analysis. Individual researchers will undoubtedly utilize different coding methods given that the researcher is the primary instrument of analysis in qualitative studies. The more a researcher shares about their coding process, the better idea a reader has of how that researcher arrived at a particular result or analysis. Navy et al. (2020) provided a rich description of their coding process guiding the reader through multiple cycles of coding, creation of codes and subcodes, and the specific kind of coding that was utilized at each stage. With this, readers can more fully confirm the article's results as they have a deeper understanding of the methodological process that guided the research. Other authors gave few details about their coding process. For example, Walan (2020) coded deductively based on teachers' use and knowledge of technology in the classroom. The author gave little information regarding their coding process beyond coding alongside a scholar knowledgeable about TPACK and that they "totally agreed on all of them without any differences in our interpretations of data" (2018, p. 434). Additionally, the author stated that "inductive themes emerged through interpretive readings of the interview transcripts" (2018, p. 434). Claims that themes emerge give agency to data, taking away or obfuscating the researcher's role in data analysis.

While there is no agreement around how much of the research process should be described in a publication, descriptive detailing of research methods is an important facet of qualitative research. King and Pringle (2019) provided an example of this level of detail of their methods:

> We used a word processing program and copied and pasted all of the passages of interest for each participant into her own running document as a single transcript and read even closer to select the most compelling pieces to start crafting the counter-stories (p. 553).

This kind of detailed explanation increases the transparency of the researchers' methodology, which helps establish their findings' credibility, dependability, and transferability.

Understandably there is an inherent dissonance between qualitative research methodology from a traditionally positivist area such as science and by association science education. Velasco et al. (2021) are an example of researchers who straddled the two paradigms of research. Despite their credible qualitative language in the methods section, they justified their inter-rater reliability with descriptive statistics as percentages in the analysis. Another example is Fitzgerald et al. (2019) who based their study on interview data but applied Bayesian confirmatory analysis to their communities of concepts, which is a quantitative statistical inference method.

Further aspects of deep engagement with qualitative research methods include positionality and reflexivity. King and Pringle's (2019) study was the only article that included "Subjectivity Statements" from each author. They prefaced with "our personal histories, cultural worldviews, and professional experiences color our lens and decisions for how we approached this study" (King & Pringle, 2019, p. 543). Strachan (2020) included a "researcher positionality" section in acknowledging their outsider status with their participants. Strachan also addressed paying particular attention to matters of race and culture so as to not solidify "racialized deficit perspectives" (2020, p. 229).

Limitations

Chang et al. (2010) and Karampelas (2021) reviewed 1,401 and 6,504 articles, respectively, which can be considered to be more representative of science education research generally. For our project, we purposefully chose to examine a smaller number of articles to enable us to more carefully examine each article's methodology. As a result, the articles selected may not be representative of the breadth of published science education research.

We were unable to account for the variable of reviewer input in the publication process. The 25 articles comprising this literature review were sourced from 16 individual journals. Each journal has countless reviewers, each possessing individual understandings and experiences with research paradigms. The anonymity given to reviewers makes it difficult to know a reviewer's evaluation criteria or paradigmatic orientations rendering this variable unable to be accounted for.

What we know about qualitative research is the result of completing a combined 48 hours of coursework in qualitative methodology at our institution as well as preparing for and presenting on original qualitative research at a combined 10 conferences, and countless hours of contact with colleagues and scholars engaged in qualitative inquiry. Through our doctoral studies, we have come to understand the history of qualitative research and its current status in our individual fields of study. We have engaged with countless methodological texts that are cited frequently in the literature and that have been written from well-established figures within qualitative research. Our analysis, while based on these experiences, which we consider to have made us well-informed on the topic, our analysis is our own and could differ from other scholars who have differing understandings of qualitative research. With this in mind, we were proactive in ensuring that we shared as much detail as possible about our analysis, coding, and other methods so that others more fully understood our analysis.

Ethical considerations in qualitative research extend far beyond approval from IRB or other institutional interests. Researchers must address a host of considerations during research planning, data collection, and reporting results. These considerations can be highly contextual based on the scope of the researcher's inquiry and their own personal researcher identity. Moreover, Roth and von Unger (2018) wrote that ethical issues can manifest in any phase of research requiring ethical reflexivity to be a core consideration of qualitative research. While we wished to analyze the 25 articles for the researchers' ethical considerations, we ultimately decided to save this inquiry for another project so that we may more fully attune to ethical considerations in research.

CONCLUSIONS

Chang et al. (2010) and Karampelas (2021) identified from their literature reviews the topics covered in science education research over previous decades. We designed our study to go beyond their work with an in depth focus on the qualitative methodologies used in secondary science teaching research. Additionally, whereas Chang et al. (2010) and Karampelas' (2021) reviewed articles from a range of years, (1990-2007 and 2010-2020, respectively), our review considered only articles published since 2018. While Chang et al. (2010) and Karampelas (2021) claimed that qualitative science education research was on the rise, findings from our small scale study revealed few articles addressed science teaching and learning with deep qualitative engagement. While several articles contained elements that represent rigorous, high-quality engagement with qualitative research practices, there was no single article included in our review that was exemplary in all of its methodology.

Education, particularly science education, has long been dominated by methodological positivism and paradigm shifts that present difficulties to researchers (Kuhn, 1970). The struggle with this dichotomy can be seen in the Velasco et al. (2021) publication as well in our own findings, where we describe qualitative research features in the articles in terms of quantitative percentages. Likewise, Dogan et al. (2020) took the process of observation, an opportunity to provide rich qualitative data, and quantified their observations with the reformed teaching observation protocol (Sawada et al., 2002). Qualitative research is new compared to quantitative research and has gone through several phases. The development to this current moment has been driven by questions related to democracy, race, gender, class, freedom, and community (Given, 2008). The present paradigm, just as paradigms of the past, does not exist as a monolith, and flexibility is inherent and arguably needed as qualitative research pushes into the future. However, this does not mean that researchers evoking qualitative methodologies should ignore the developments in qualitative research, which have defined it over the years.

Author contributions: All authors were involved in concept, design, collection of data, interpretation, writing, and critically revising the article. All authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Acknowledgements: The authors would like to thank Dr. Stephanie Anne Shelton for the mentorship.

Ethics declaration: The authors declared that ethics committee approval was not required for this study as it was based on existing literature.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

Abell, S. K. & Lederman, N. G. (2007). *Handbook of research on science education*. Lawrence Erlbaum Associates.

Anderson, V. (2017). Criteria for evaluating qualitative research. *Human Resource Development Quarterly, 28*(2), 125-133. https://doi.org/10.1002/hrdq.21282

Andrée, M., & Hansson, L. (2021). Industry, science education, and teacher agency: A discourse analysis of teachers' evaluations of industry-produced teaching resources. *Science Education, 105*, 353-383. https://doi.org/10.1002/sce.21607

Avraamidou, L. (2014). Studying science teacher identity: Current insights and future research directions. *Studies in Science Education, 50*(2), 145-179. https://doi.org/10.1080/03057267.2014.937171

Berge, M., Danielsson, A., & Lidar, M. (2020). Storylines in the physics teaching content of an upper secondary school classroom. *Research in Science & Technological Education, 38*(1), 63-83. https://doi.org/10.1080/02635143.2019.1593128

Birth, M., Claes, D. R., & Pedersen, J. E. (2018). Physics teachers as physics experts: Research participation as professional development. *Science Educator, 26*(2), 90-101.

Brown, R. E., & Bogiages, C. A. (2019). Professional development through STEM integration: How early career math and science teachers respond to experiencing integrated STEM tasks. *International Journal of Science and Mathematics Education, 17*(1), 111-128. https://doi.org/10.1007/s10763-017-9863-x

Chang, Y. H., Chang, C. Y., & Tseng, Y. H. (2010). Trends of science education research: An automatic content analysis. *Journal of Science Education and Technology, 19*, 315–331. https://doi.org/10.1007/s10956-009-9202-2

Cooper, H. (2017). *Research synthesis and meta-analysis: A Step-by-step approach*. SAGE. https://doi.org/10.4135/9781071878644

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. SAGE. https://doi.org/10.4135/9781452230153

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* SAGE.

Creswell, J., & Plano Clark, V. (2011). *Designing and conducting mixed methods research*. SAGE.

Dogan, O. K., Cakir, M., Tillotson, J. W., Young, M., & Yager, R. E. (2020). A longitudinal study of a new science teacher's beliefs and classroom practices. *International Journal of Progressive Education, 16*(1), 84-99. https://doi.org/10.29329/ijpe.2020.228.7

Dolfing, R., Prins, G. T., Bulte, A. M. W., Pilot, A., & Vermunt, J. D. (2020). Strategies to support teachers' professional development regarding sense-making in context-based science curricula. *Science Education, 105*, 127-165. https://doi.org/10.1002/sce.21603

Duit, R. (2007). Science education research internationally: Conceptions, research methods, domains of research. *Eurasia Journal of Mathematics, Science and Technology Education, 3*(1), 3-15. https://doi.org/10.12973/ejmste/75369

Fitzgerald, M., Danaia, L., & McKinnon, D. H. (2019). Barriers inhibiting inquiry-based science teaching and potential solutions: Perceptions of positively inclined early adopters. *Research in Science Education, 49*, 543-566. https://doi.org/10.1007/s11165-017-9623-5

Flick, U. (2022). Revitalizing triangulation for designing multi-perspective qualitative research. In U. Flick (Ed.), *The SAGE handbook of qualitative research design* (pp. 652-664). SAGE. https://doi.org/10.4135/9781529770278

Gardner, M., & Tillotson, J. W. (2019). Interpreting integrated STEM: Sustaining pedagogical innovation within a public middle school context. *International Journal of Science and Mathematics Education, 17*, 1283-1300. https://doi.org/10.1007/s10763-018-9927-6

Given, L. M. (2008). *The SAGE encyclopedia of qualitative research methods*. SAGE. https://doi.org/10.4135/9781412963909

Groenewald, T. (2004). A phenomenological research design illustrated. *International Journal of Qualitative Methods, 3*(1), 42-55. https://doi.org/10.1177/160940690400300104

Hordvik, M., Fletcher, T., Haugen, A. L., Moller, L., & Engebretsen, B. (2021). Using collaborative self-study and rhizomatic to explore the ongoing nature of becoming teacher educators. *Teaching and Teacher Education, 101*, 103318. https://doi.org/10.1016/j.tate.2021.103318

Jensen, D. (2008a). Confirmability. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 112-113). SAGE. https://doi.org/10.4135/9781412963909.n60

Jensen, D. (2008b). Credibility. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 139). SAGE. https://doi.org/10.4135/9781412963909.n77

Jensen, D. (2008c). Dependability. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 209). SAGE. https://doi.org/10.4135/9781412963909.n106

Jensen, D. (2008d). Transferability. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 886-887). SAGE. https://doi.org/10.4135/9781412963909.n464

Karampelas, K. (2021). Trends on science education research topics in education journals. *European Journal of Science and Mathematics Education, 9*(1), 1-12. https://doi.org/10.30935/scimath/9556

Keiler, L. S. (2018). Teachers' roles and identities in student-centered classrooms. *International Journal of STEM Education, 5*, 34. https://doi.org/10.1186/s40594-018-0131-6

King, N. S., & Pringle, R. M. (2019). Black girls speak STEM: Counter-stories of informal and formal learning experiences. *Journal of Research in Science Teaching, 56*, 539-569. https://doi.org/10.1002/tea.21513

Kirmaci, M., Buxton, C. A., & Allexsaht-Snider, M. (2019). Being on the other side of the table: A qualitative study of a community-based science learning program with Latinx families. *Urban Education, 58*(4), 675-707. https://doi.org/10.1177/0042085919877934

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice, 24(*1), 120-124. https://doi.org/10.1080/13814788.2017.1375092

Kuhn, T. J. (1970). *The structure of scientific revolutions*. University of Chicago Press.

Lederman, N. G., & Abell, S. K. (2014). *Handbook of research on science education*. Routledge. https://doi.org/10.4324/9780203097267

Libarkin, J. C., & Kurdziel, J. P. (2002). Research methodologies in science education: The qualitative-quantitative debate. *Journal of Geoscience Education, 50*(1), 78-86. https://doi.org/10.1080/10899995.2002.12028053

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE. https://doi.org/10.1016/0147-1767(85)90062-8

Lee, M.-H., Wu, Y.-T., & Tsai, C.-C. (2009). Research trends in science education from 2003 to 2007: A content analysis of publications in selected journals. *International Journal of Science Education, 31*(15), 1999-2020. https://doi.org/10.1080/09500690802314876

Lin, T.-J., Lin, T.-C., & Tsai, C.-C. (2013). Research trends in science education from 2008 to 2012: A systematic content analysis of publications in selected journals. *International Journal of Science Education, 36*(8), 1346-1372. https://doi.org/10.1080/09500693.2013.864428

Lin, T.-J., Lin, T.-C., Potvin, P., & Tsai, C.-C. (2018). Research trends in science education from 2013 to 2017: A systematic content analysis of publications in selected journals. *International Journal of Science Education, 41*(3), 367-387. https://doi.org/10.1080/09500693.2018.1550274

Litman, C., & Greenleaf, C. (2018). Argumentation tasks in secondary English language arts, history, and science: Variations in instructional focus and inquiry space. *Reading Research Quarterly, 53*(1), 107-126. https://doi.org/10.1002/rrq.187

Lundqvist, E., & Sund, P. (2018). Selective traditions in group discussions: Teachers' views about good science and the possible obstacles when encountering a new topic. *Cultural Studies of Science Education, 13*(2), 353-370. https://doi.org/10.1007/s11422-016-9768-y

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach*. SAGE.

Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. Jossey-Bass.

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. Jossey-Bass.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. Jossey-Bass.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Navy, S. L., Nixon, R. S., Luft, J. A., & Jurkiewicz, M. A. (2020). Accessed or latent resources? Exploring new secondary science teachers' networks of resources. *Journal of Research in Science Teaching, 57*, 184-208. https://doi.org/10.1002/tea.21591

Nixon, R. S., Toerien, R., & Luft, J. A. (2019). Knowing more than their students: Characterizing secondary science teachers' subject matter knowledge. *School Science and Mathematics, 119*, 150-160. https://doi.org/10.1111/ssm.12323

NRC. 2012. *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

NSTA. (2017). Position statement: The role of research on science teaching and learning. *National Science Teacher Association*. https://www.nsta.org/nstas-official-positions/role-research-science-teaching-and-learning

Overman, M., Vermunt, J. D., Meijer, P. C., & Brekelmans, M. (2019). Teacher-student negotiations during context-based chemistry reform: A case study. *Journal of Research in Science Teaching, 56*, 797-820. https://doi.org/10.1002/tea.21528

Poland, B. D. (2011). Transcription quality. In J. F. Gubrium, & J. A. Holstein (Eds.), *Handbook of interview research* (pp. 628-649). SAGE. https://doi.org/10.4135/9781412973588.d36

Pringle, R. M., Dawson, K., & Ritzhaupt, A. D. (2015). Integrating science and technology: Using technological pedagogical content knowledge as a framework to study the practices of science teachers. *Journal of Science Education and Technology, 24*(5), 648-662. https://doi.org/10.1007/s10956-015-9553-9

Robertson, W. B., & Yazan, B. (2022). Navigating tensions and asserting agency in language teacher identity: A case study of a graduate teaching assistant. *Linguistics and Education, 71*, 101079. https://doi/org.10.1016/j.linged.2022.101079

Roth, W.-M., & von Unger, H. (2018). Current perspectives on research ethics in qualitative research. *Forum: Qualitative Social Research, 19*(3), 33. https://doi.org/10.17169/fqs-19.3.3155

Saldaña, J. (2016). *The coding manual for qualitative researchers*. SAGE.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics, 102*(6), 245-253. https://doi.org/10.1111/j.1949-8594.2002.tb17883.x

Smith, L. (1979). An evolving logic of participant observation, educational ethnography, and other case studies. In L. Shulman (Ed.), *Review of research in education* (pp. 316-277). F. E. Peacock. https://doi.org/10.2307/1167249

Stake, R. E. (1995). *The art of case study research*. SAGE.

Steinke, I. (2004). Quality criteria in qualitative research. In W. Flick, E. von Kardoff, & I. Steinke (Eds.) *A companion to qualitative research* (pp. 184-190). SAGE.

Strachan, S. L. (2020). An examination of two African American males' decisions to become secondary science teachers. *The High School Journal, 103*(4), 221-240. https://doi.org/10.1353/hsj.2020.0014

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837-851. https://doi.org/10.1177/1077800410383121

Tsai, C.-C., & Wen, M. L. (2011). Research and trends in science education from 1998 to 2002: A content analysis of publication in selected journals. *The International Journal of Science Education, 27*(3), 3-14. https://doi.org/10.1080/0950069042000243727

Vagle, M. D. (2014). *Crafting phenomenological research*. Left Coast Press, Inc.

Vale, C., Campbell, C., Speldewinde, C., & White, P. (2020). Teaching across subject boundaries in STEM: Continuities in beliefs about learning and teaching. *International Journal of Science and Mathematics Education, 18*(3), 463-483. https://doi.org/10.1007/s10763-019-09983-2

Vázquez-Bernal, B., Mellado, V., & Jiménez-Pérez, R. (2021). The long road to shared PCK: A science teacher's personal journey. *Research in Science Education, 52*, 1807-1828. https://doi.org/10.1007/s11165-021-10028-4

Velasco, R. C. L., Hite, R., & Milbourne, J. (2021). Exploring advocacy self-efficacy among K-12 STEM teacher leader. *International Journal of Science and Mathematics Education, 20*, 435-457. https://doi.org/10.1007/s10763-021-10176-z

Verbatim. (n. d.). Merriam-Webster.com dictionary. *Merriam Webster*. https://www.merriam-webster.com/dictionary/verbatim

Verbatim, adv., adj., and n. (2023) *OED online*. Oxford University Press.

Vossen, T. E., Henze, I., De Vries, M. J., & Van Driel, J. H. (2020). Finding the connection between research and design: The knowledge development of STEM teachers in a professional learning community. *International Journal of Technology and Design Education, 30*, 295-320. https://doi.org/10.1007/s10798-019-09507-7

Walan, S. (2020). Embracing digital technology in science classrooms–Secondary school teachers' enacted teaching and reflections on practice. *Journal of Science Education and Technology, 29*, 431-441. https://doi.org/10.1007/s10956-020-09828-6

Wen, Y., Wu, L., & He, S. (2021). Investigating affordances and tensions in STEM applied learning program from practitioners' sensemaking. *International Forum of Educational Technology & Society, 24*(4), 99-109.

Wilson, K. (2021). Exploring the challenges and enablers of implementing a STEM project-based learning program in a diverse junior secondary context. *International Journal of Science and Mathematics Education, 19*(5), 881-897. https://doi.org/10.1007/s10763-020-10103-8

Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam and Stake. *The Qualitative Report, 20*(2), 134-152. https://doi.org/10.46743/2160-3715/2015.2102

APPENDIX A

Table A1. Articles

Reference	Title	Publication place	DOI
Andrée and Hansson (2021)	Industry, science education, and teacher agency: A discourse analysis of teachers' evaluations of industry-produced teaching resources	*Science Education, 105*, 353-383	https://doi.org/10.1002/hrdq.21282
Berge et al. (2020)	Storylines in the physics teaching content of an upper secondary school classroom	*Research in Science & Technological Education, 38*(1), 63-83	https://doi.org/10.1080/02635143.2019.1593128
Birth et al. (2018)	Physics teachers as physics experts: Research participation as professional development	*Science Educator, 26*(2), 90-101	
Brown and (2019)	Professional development through STEM integration: How early career math and science teachers respond to experiencing integrated STEM tasks	*International Journal of Science and Mathematics Education, 17*(1), 111-128	https://doi.org/10.1007/s10763-017-9863-x
Dogan et al. (2020)	A longitudinal study of a new science teacher's beliefs and classroom practices	*International Journal of Progressive Education, 16*(1), 84-99	https://doi.org/10.29329/ijpe.2020.228.7
Dolfing et al. (2020)	Strategies to support teachers' professional development regarding sense-making in context-based science curricula	*Science Education, 105*, 127-165	https://doi.org/10.1002/sce.21603
Fitzgerald et al. (2019)	Barriers inhibiting inquiry-based science teaching and potential solutions: Perceptions of positively inclined early adopters	*Research in Science Education, 49*, 543-566	https://doi.org/10.1007/s11165-017-9623-5
Gardner and (2019)	Interpreting integrated STEM: Sustaining pedagogical innovation within a public middle school context	*International Journal of Science and Mathematics Education, 17*, 1283-1300	https://doi.org/10.1007/s10763-018-9927-6
Hordvik et al. (2021)	Using collaborative self-study and rhizomatic to explore the ongoing nature of becoming teacher educators	*Teaching and Teacher Education, 101*, 103318	https://doi.org/10.1016/j.tate.2021.103318
Keiler (2018)	Teachers' roles and identities in student-centered classrooms	*International Journal of STEM Education, 5*, 34	https://doi.org/10.1186/s40594-018-0131-6
King and (2019)	Black girls speak STEM: Counter-stories of informal and formal learning experiences	*Journal of Research in Science Teaching, 56*, 539-569	https://doi.org/10.1002/tea.21513
Kirmaci et al. (2019)	Being on the other side of the table: A qualitative study of a community-based science learning program with Latinx families	*Urban Education, 58*(4), 675-707	https://doi.org/10.1177/0042085919877934
Litman and (2018)	Argumentation tasks in secondary English language arts, history, and science: Variations in instructional focus and inquiry space	*Reading Research Quarterly, 53*(1), 107-126	https://doi.org/10.1002/rrq.187
Lundqvist and (2018)	Selective traditions in group discussions: Teachers' views about good science and the possible obstacles when encountering a new topic	*Cultural Studies of Science Education, 13*(2), 353-370	https://doi.org/10.1007/s11422-016-9768-y
Navy et al. (2020)	Accessed or latent resources? Exploring new secondary science teachers' networks of resources	*Journal of Research in Science Teaching, 57*, 184-208	https://doi.org/10.1002/tea.21591
Nixon et al. (2019)	Knowing more than their students: Characterizing secondary science teachers' subject matter knowledge	*School Science and Mathematics, 119*, 150-160	https://doi.org/10.1111/ssm.12323
Overman et al. (2019)	Teacher-student negotiations during context-based chemistry reform: A case study	*Journal of Research in Science Teaching, 56*, 797-820	https://doi.org/10.1002/tea.21528
Strachan (2020)	An examination of two African American males' decisions to become secondary science teachers	*The High School Journal, 103*(4), 221-240	https://doi.org/10.1353/hsj.2020.0014
Vale et al. (2020)	Teaching across subject boundaries in STEM: Continuities in beliefs about learning and teaching	*International Journal of Science and Mathematics Education, 18*(3), 463-483	https://doi.org/10.1007/s10763-019-09983-2
Vázquez-Bernal et al. (2021)	The long road to shared PCK: A science teacher's personal journey	*Research in Science Education, 52*, 1807-1828	https://doi.org/10.1007/s11165-021-10028-4

Table A1 (Continued). Articles

Reference	Title	Publication place	DOI
Velasco et al. (2021)	Exploring advocacy self-efficacy among K-12 STEM teacher leader	*International Journal of Science and Mathematics Education, 20*, 435-457	https://doi.org/10.1007/s10763-021-10176-z
Vossen et al. (2020)	Finding the connection between research and design: The knowledge development of STEM teachers in a professional learning community	*International Journal of Technology and Design Education, 30*, 295-320	https://doi.org/10.1007/s10798-019-09507-7
Walan (2020)	Embracing digital technology in science classrooms–Secondary school teachers' enacted teaching and reflections on practice	*Journal of Science Education and Technology, 29*, 431-441	https://doi.org/10.1007/s10956-020-09828-6
Wen et al. (2021)	Investigating affordances and tensions in STEM applied learning program from practitioners' sensemaking	*International Forum of Educational Technology & Society, 24*(4), 99-109	
Wilson (2021)	Exploring the challenges and enablers of implementing a STEM project-based learning program in a diverse junior secondary context	*International Journal of Science and Mathematics Education, 19*(5), 881-897	https://doi.org/10.1007/s10763-020-10103-8

APPENDIX B

Table B1. Variables & descriptions

Variable	Description
Role of researcher	Was researcher a participant or observer (i.e., P: Participant & O: Observer)?
Methodology	Did the author explicitly state a methodology? If so, what (e.g., case study, narrative/discourse inquiry, phenomenology, ethnography, etc.)?
Data collection	What kinds of data collection did the author explicitly state in the article (e.g., interview, observation, journal, fieldnotes, & focus group)?
Time durations	Did the author explicitly state the duration of time spent in the field? If so, how long?
Triangulation	Did the author explicitly detail how data collected was triangulated? If so, how (e.g., data, method, & researcher triangulation for case study in particular)?
Member-checking	Did the author explicitly state that member checks were carried out? If so, how (e.g., participants were given voice/choice)?
Coding method	Did the author explicitly state the approach to coding their collected data? If so, what approach (e.g., NVivo, manual, narrative analysis, inductive, deductive, etc.)?

APPENDIX C

Table C1. Articles

Citation	Role of researcher (P= Participant; O= Observer)	Methodology	Data Collection	Time durations (Y=yes; N=no)	Triangulation (Y-yes; N=no)	Member check (Y=yes; N=no)	Coding
Andrée, M., & Hansson, L. (2021). Industry, science education, and teacher agency: A discourse analysis of teachers' evaluations of industry-produced teaching resources. *Science Education, 105*, 353-383. https://doi.org/10.1002/sce.21607	O	Discourse Analysis	5 Focus Groups	Y; 1 hr focus group	N	N	Deductive - Biesta and Tedder's (2006) ecological model of teacher agency
Berge, M., Danielsson, A., & Lidar, M. (2020). Storylines in the physics teaching content of an upper secondary school classroom. *Research in Science & Technological Education, 38*(1), 63-83. https://doi.org/10.1080/02635143.2019.1593128	O	Case Study	Classroom observation, interviews, 1st analysis discourse; 2nd analysis of videos for body language	Y; 21–25 min videos	N	N	Deductive - coded speech acts based on 3 analytical questions
Birth, M., Claes, D. R., & Pedersen, J. E. (2018). Physics teachers as physics experts: Research participation as professional development. *Science Educator, 26*(2), 90-101.	O	Phenomenology	interviews	N	N	N	Inductive - "Emergent themes"
Brown, R. E., & Bogiages, C. A. (2019). Professional development through STEM integration: How early career math and science teachers respond to experiencing integrated STEM tasks. *International Journal of Science and Mathematics Education, 17*(1), 111-128. https://doi.org/10.1007/s10763-017-9863-x	O	Case Study	written reflections	N	N	N	Deductive - constant comparative analysis, a prioi codes
Dogan, O. K., Cakir, M., Tillotson, J. W., Young, M., & Yager, R. E. (2020). A longitudinal study of a new science teacher's beliefs and classroom practices. *International Journal of Progressive Education, 16*(1), 84-99. https://doi.org/10.29329/ijpe.2020.228.7	O	Case Study	interviews, questionnaires, classroom observations	Y; 1-1.5 hr interviews	Y; data - interview, survey, observation	N	Both - 4 dimensions methodological, conceptual, social, & epistemological (Furtak 2006), themes emerged
Dolfing, R., Prins, G. T., Bulte, A. M. W., Pilot, A., & Vermunt, J. D. (2020). Strategies to support teachers' professional development regarding sense-making in context-based science curricula. *Science Education, 105*, 127-165. https://doi.org/10.1002/sce.21603	O	Case Study	Not explicit - "focus groups" "data sources were collected", audio recordings, video recordings, written	N	Y	Intentionally No	Deductive - coded based on aspects on teaching science curricula (fig 2)
Fitzgerald, M., Danaia, L., & McKinnon, D. H. (2019). Barriers inhibiting inquiry-based science teaching and potential solutions: Perceptions of positively inclined early adopters. *Research in Science Education, 49*, 543–566. https://doi.org/10.1007/s11165-017-9623-5	O	Not declared (qualitative)	interviews	Y; 40-120 min interviews	N	N	Inductive - emerging themes, Leximancer
Gardner, M., & Tillotson, J. W. (2019). Interpreting integrated STEM: Sustaining pedagogical innovation within a public middle school context. *International Journal of Science and Mathematics Education, 17*, 1283-1300. https://doi.org/10.1007/s10763-018-9927-6	O	Phenomenology; descriptive case study	observations, interviews, pre- & post- lesson feedback (informal interviews)	Y; recorded 1383 min. instruction	N	Y; extended engagement with participants and member-checking throughout (not post	Inductive - thematic collation, Atlas Ti

Table C1 (continued). Articles

Citation	Role of researcher (P= Participant; O= Observer)	Methodology	Data Collection	Time durations (Y=yes; N=no)	Triangulation (Y-yes; N=no)	Member check (Y=yes; N=no)	Coding
Hordvik, M., Fletcher, T., Haugen, A. L., Moller, L., & Engebretsen, B. (2021) Using collaborative self-study and rhizomatics to explore the ongoing nature of becoming teacher educators. *Teaching and Teacher Education 101*, 1-11. https://doi.org/10.1016/j.tate.2021.103318	P	dialogic S-STEP methodology & rhizomatic thinking	audio recordings, reflective diaries	Y; 35 hour audio recording	N	Y; participant as critical friend, co-authorship, reviewed researchers' reflections, met for discussions	Inductive - interactive analytic process
Keiler, L. S. (2018). Teachers' roles and identities in student-centered classrooms. *International Journal of STEM Education 5*(34), 1-20. https://doi.org/10.1186/s40594-018-0131-6	P	Case Study	interviews, focus groups	Y; 20-90 min focus groups	Y; data - interview, focus group, other data sources	Y; participant written feedback with revisions	Inductive - iterative process, emerging codes
King, N. S., & Pringle, R. M. (2019). Black girls speak STEM: Counterstories of informal and formal learning experiences. *Journal of Research in Science Teaching, 56*, 539-569. https://doi.org/10.1002/tea.21513	P	Narrative inquiry	interviews, journals, student samples, researcher memos, observations	Y; 60 min interview	Y; interview, participant reflective journal, student work samples, researcher memo	Y co-researchers co-construction	Inductive - emerging themes; NVivo
Kirmaci, M., Buxton, C. A., & Allexsaht-Snider, M. (2019). Being on the other side of the table: A qualitative study of a community-based science learning program with Latinx families. *Urban Education*, 1-33. https://doi.org/10.1177/0042085919877934	O	multi-case study	interview, observation, focus group	N	Y; interview, focus group, observation	Y; verify analysis with participants	Inductive - Charmaz's constant comparative analysis & emerging themes; ATLAS.ti 8.0
Litman, C., & Greenleaf, C. (2018). Argumentation tasks in secondary English language arts, history, and science: Variations in instructional focus and inquiry space. *Reading Research Quarterly, 53*(1), 107-126. https://doi.org/10.1002/rrq.187	O	not explicitly stated	observations, fieldnotes, audio recording, video recording, artifacts, interviews, interpretive summary	Y; 34.4 video recording	N	Y; verify observations with participants	Both - combination of inductive and theoretically driven analyse
Lundqvist, E., & Sund, P. (2018). Selective traditions in group discussions: teachers' views about good science and the possible obstacles when encountering a new topic. *Cultural Studies of Science Education, 13*(2), 353-370. https://doi.org/10.1007/s11422-016-9768-y	O	Group Conversation Analysis	3 focus group sessions	Y; 75 min group interview	N	N	Deductive - not explicitly stated
Navy, S. L., Nixon, R. S., Luft, J. A., & Jurkiewicz, M. A. (2020). Accessed or latent resources? Exploring new secondary science teachers' networks of resources. *Journal of Research in Science Teaching. 57*, 184-208. https://doi.org/10.1002/tea.21591	O	Observational study	interview, observation	Y; 45-60 min interview	Y; investigator triangulation, data, triangulation	N	Both - a pirori & inductive; Saldana's 1st & 2nd cycle coding method; NVivo 11
Nixon, R. S., Toerien, R., & Luft, J. A. (2019). Knowing more than their students: Characterizing secondary science teachers' subject matter knowledge. *School Science and Mathematics, 119*, 150-160. https://doi.org/10.1111/ssm.12323	O	not specified	interview	Y; 20 min interview	N	N	Deductive - coding for specialized content knowledge
Overman, M., Vermunt, J. D., Meijer, P. C., Brekelmans, M. (2019). Teacher–student negotiations during context-based chemistry reform: A case study. *Journal of Research in Science Teaching, 56*, 797-820. https://doi.org/10.1002/tea.21528	O	Case Study	questionnaire, interview, video-stimulated interviews	Y; 1.5-2 hour interviews	N	Y; participant reflections on video recordings	Inductive - grounded theory, themes emerged

Table C1 (continued). Articles

Citation	Role of researcher (P= Participant; O= Observer)	Methodology	Data Collection	Time durations (Y=yes; N=no)	Triangulation (Y-yes; N=no)	Member check (Y=yes; N=no)	Coding
Strachan, S. L. (2020). An examination of two African American males' decisions to become secondary science teachers. *The High School Journal, 103*(4), 221-240. https://doi.org/10.1353/hsj.2020.0014	O	descriptive phenomenology	interview	Y; 1 hour interviews	N	Y; data shared, follow-up conversation	Inductive - 2 level, open coding
Vale, C., Campbell, C., Speldewinde, C., & White, P. (2020). Teaching across subject boundaries in STEM: Continuities in beliefs about learning and teaching. *International Journal of Science and Mathematics Education, 18*(3), 463-483. https://doi.org/10.1007/s10763-019-09983-2	O	not specified	interview, video-stimulated interviews	N	N	N	Deductive - constant comparative analysis; NVivo
Vázquez-Bernal, B., Mellado, V., & Jiménez-Pérez, R. (2021). The long road to shared PCK: A science teacher's personal journey. *Research in Science Education*. https://doi.org/10.1007/s11165-021-10028-4	O	longitudinal case study	ethnographic records = diary, questionnaire, interview, observations	Y; recorded in table	Y; ethnographic records, participant reflection, observation	Y; participant elaborated researchers reports with narrative	Deductive - methodological plurality of reflections and actions; AQUAD
Velasco, R.C.L., Hite, R., & Milbourne, J. (2021). Exploring advocacy self-efficacy among K-12 STEM teacher leader. *International Journal of Science and Mathematics Education, 20*, 435-457. https://doi.org/10.1007/s10763-021-10176-z	P	Case Study	interview, focus group, documentation	Y; 45 min interview, 45 min focus group	Y; interview, focus group, documentation	Implied as insider-researcher	Deductive - Self-efficacy lens, Saldana 2-coding cycle; NVivo 12
Vossen, T. E., Henze, I., De Vries, M. J., & Van Driel, J. H. (2020). Finding the connection between research and design: The knowledge development of STEM teachers in a professional learning community. *International Journal of Technology and Design Education, 30*, 295-320. https://doi.org/10.1007/s10798-019-09507-7	O	multi-case study	interview, PLC meeting recordings, CoRe (PCK self-eval)	N	N	N	Deductive - 4 domains of PCK (Magnusson et al); Atlas.ti
Walan, S. (2020). Embracing digital technology in science classrooms - Secondary school teachers' enacted teaching and reflections on practice. *Journal of Science Education and Technology, 29*, 431-441. https://doi.org/10.1007/s10956-020-09828-6	O	Case Study	observation, interview before and after lesson, follow-up interviews	Y; 9.5 hours observation, 9 hours interviews	Y; interviews & observation, researcher triangulation	Y; participants reviewed fieldnotes	Deductive - coded for TK, TPK, TCK, TPACK categories
Wen, Y., Wu, L., & He, S. (2021). Investigating affordances and tensions in STEM applied learning programme from practitioners' sensemaking. *International Forum of Educational Technology & Society, 24*(4), 99-109.	O	Case Study	interview, focus group,	Y; 50-57 min interviews	N	N	Inductive - emerging themes
Wilson, K. (2021). Exploring the challenges and enablers of implementing a STEM project-based learning programme in a diverse junior secondary context. *International Journal of Science and Mathematics Education, 19*(5), 881-897. https://doi.org/10.1007/s10763-020-10103-8	O	Case Study	interview, observation, curriculum documents	Y; 60 min interviews, 1960 min observations	Y; observation, interview, documentation	Y; all data verified with participants	Inductive - emerging themes; NVivo 10