ADETOGUN, ADEYEMO A. Ph.D. Discourse of STEM Education Evaluation: Current and Future Perspectives. (2023)
Directed by Dr. Ayesha Boyce & Dr. Aileen Reid. 241 pp.

Science, Technology, Engineering, and Mathematics (STEM) education has become increasingly important in the US due to its influence on the nation's educational needs, the creation of a skilled labor force, and opportunities for more tech-savvy workers. However, the evaluation approaches and methodologies used in STEM education programs have come under examination in recent times, with questions being raised in the evaluation literature. Evaluation practitioners in STEM education come from diverse backgrounds, hold different worldviews, and subscribe to a growing list of evaluation approaches. These practitioners value quantitative, qualitative, or mixed methods in evaluation objectives, questions posed, study design, implementation, report writing, and dissemination.

This dissertation is sequential qualitative research that examined the Discourse of STEM education evaluation from a practitioner's perspective. It focused on the dominant evaluation approaches and methodologies while reviewing AEA STEM TIG abstracts and conducting individual interviews with STEM education evaluators. The research revealed the varied landscape of evaluation approaches stemming from differing values, training, and interests of evaluators. The study also illuminated the philosophical underpinnings of method choices that shape evaluation methodologies.

Several findings emerged from this research. These include STEM education evaluators' discussion on lessons learned, challenges, and impact on STEM education evaluation practice. Furthermore, collaborative evaluation approaches were frequently used to understand the complexity of STEM education programs, and mixed methodologies were used to strengthen

studies by incorporating both qualitative and quantitative approaches. STEM education evaluators from academic, private, and public sectors discussed program planning, implementation, and professional development for educators, as well as the use of technology in practice. Differences emerged, with academic evaluators focusing on cultural context, undergraduate research mentorship, and evaluation instruments, private evaluators emphasizing fidelity of implementation, data collection, and collective impact measurements, and public evaluators caring about program culture, career plans, work-life balance, and workforce trajectories.

STEM education evaluators emphasized the importance of equity in broadening participation, spoke positively about professional development and community organizing, noted the evolution of accountability through leadership, social justice, and self-socialization, recognized the significant influence of new technology on data utilization in evaluation, and discussed new approaches and methodologies to advance STEM education evaluation. The STEM evaluation literature identifies various evaluation approaches that resonate with STEM education evaluators. Evaluators in the STEM fields are mostly interested in Client-Centered/Responsive and Utilization-Focused Evaluation Approaches. Academic and private sector evaluators share common ground in Collaborative/Partnership Approaches, Culturally Responsive Evaluation, Outcome/Value-Added Assessment, and Decision/Accountability. In addition, evaluators in the STEM fields justify their evaluation methodologies with data collection devices, status, and surveys.

In conclusion, STEM education evaluation approaches are constantly evolving and influenced by various contexts. Evaluators from different sectors have different viewpoints on the current and future direction of the field. However, STEM education evaluators emphasize the

importance of equity in broadening participation, professional development, community organizing, and the use of technology in practice. Further exploration of the Discourse around STEM education evaluation methodologies is needed to clarify the philosophical assumptions underlying evaluation methods.

DISCOURSE OF STEM EDUCATION EVALUATION: CURRENT AND FUTURE PERSPECTIVES

by

Adetogun Adeyemo Adekanmi

A Dissertation
Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro

2023

Approved by

Dr. Ayesha Boyce
Committee Co-Chair

Dr. Aileen Reid
Committee Co-Chair

DEDICATION

This work is, foremost, dedicated to the Glory of God who is the author and finisher of my faith. In Him rest my hopes and my dreams, and through Him, I am more than a conqueror. I can soar and accomplish great things through Christ. Thank you, Lord.

I would like to express my deepest gratitude to my dearest wife, Tiara, and awesome children, Nathanael, Tejumade, and Nehemiah, who has been my pillar of support throughout this journey. Your unwavering love, encouragement, and sacrifice have made it possible for me to fulfill this dream.

To my loving mother, Mary Adunni, I am forever grateful for your unconditional love and unceasing prayers for me in this journey and always. To my dearest sister, Folasade, and dependable brother, Adeboye, thank you for your support, encouragement, and belief in me always.

APPROVAL PAGE

This dissertation written by Adeyemo Adekanmi Adetogun has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Co-Chair	Dr. Ayesha Boyce
Committee Co-Chair	Dr. Aileen Reid
Committee Members	Dr. Devdass Sunnassee
	Dr. Cherie Avent

March 17, 2023
Date of Acceptance by Committee

March 2, 2023
Date of Final Oral Examination

ACKNOWLEDGEMENTS

I want to express my special thanks to my doctoral committee, Doctors Ayesha Boyce, Devdass Sunnassee, Cherie Avent, and Aileen Reid. My deepest gratitude goes to Dr. Boyce, who doubles as my academic advisor in addition to being my doctoral committee chair. She has been there for me since day one of my graduate studies at the educational research methodology (ERM) department at the University of North Carolina Greensboro (UNCG). Through the vicissitude of the seasons, the good times, and the not-so-good ones, you helped me, supported me, encouraged me, motivated me, and your gentle albeit firm leadership and mentorship are unequaled, and I will forever be grateful for all you did for me.

I would like to acknowledge and thank some of my good friends who contributed in one way or another to make this work a reality. My good friends Olanrewaju Sanusi, Kolawole Omomowo, Mawuena Logan, Jen Gathings, Kossi Agbeve, Michael LaClair, Richard LaBennett, and many others are too numerous to mention. I thank you all for being my friend, for coming through for me time and again, for critical friendship when I needed it, for encouragement in shaky times, and support throughout this journey.

Finally, I know with God all things are possible, and I once again thank everyone who has contributed to my success in completing this dissertation. I could not have done it without your support, and I am forever grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Human and material investments in science, technology, engineering, and mathematics (STEM) education have grown exponentially since the turn of the millennium. In addition, there are growing conversations at STEM education meetings and conferences surrounding issues of access, Broadening Participation, and the inclusion of underrepresented and marginalized populations. More recently, national debates and dialogues are gaining momentum for information technology-based economy, food security, homeland, and national security, and questions are being raised about the ability of the United States (U.S.) to compete economically with the rest of the world (Anon, 2013). As these issues continue to dominate the national stage, they invariably form the core of major evaluation Discourses taking place in formal and informal meetings across the country. Notwithstanding the massive effort by the federal government, its funding agencies, and other interested stakeholders, little is known about the evaluation of STEM education programs. The current state of STEM education programs requires clarity in terms of evaluation effort, otherwise, it runs the risk of being diminished to mere buzzwords (Whitehurst, 2009).

As STEM education initiatives continue to proliferate, there is insufficient evidence regarding the efficacy of their efforts (Whitehurst, 2009). Opportunities for knowledge and skill acquisition needed to compete in a technology-dependent future are narrow (Huffman & Lawrenz, 2006), research experience programs lack clear pathways to addressing major education issues, and investments in STEM education are not matched by clear commitments to evaluation activities (Greene et al., 2006). A report by the Academic Competitiveness Council (2007) suggested discussions are ongoing at different levels of government among policymakers, in academia within evaluation theorists and scholars, and at meetings where a major confluence

of these stakeholders come together with practitioners to engage one another in current and future issues in STEM education evaluation (Whitehurst, 2009). The apparent gaps in the evaluation of STEM education initiatives have led to calls for the U.S. government, departments, agencies, and organizations with an interest in STEM education to ramp up rigorous evaluation efforts (Gonzalez, 2012).

Background

The Nucleus of STEM Education

The core of STEM enterprise in today's 21st century contains scientific thinking, groundbreaking research leading to inventions, and technological innovations designed to transform human experiences. In labor terms, the United States (U.S.) workforce continues to open new opportunities for local and global competitiveness through reliance on technological advances prompted by sound engineering systems (Chesky, & Wolfmeyer, 2015; James & Singer, 2016; National Science Foundation, 2008). Furthermore, the U.S. STEM enterprises create products and offer services deemed critical to the thriving and sustenance of its economy and have one of the highest average wages (average of $77,570 across all occupations) compared to other fields. In 2010, there were 7.6 million STEM workers in the U. S. or about 1 in every 18 U.S. workers holding a job in STEM. Employment growth projections between 2010-2020 in the STEM field were at 18.7% compared to 14.3% for all other occupations (U.S. Bureau of Labor Statistics, 2015). Relatedly, although the long-term, systemic impact of the covid-19 pandemic has yet to be established; however, unemployment rates for STEM workers were much lower than non-STEM workers between the first and second quarters of 2020 and remained high relative to pre-pandemic levels (Okrent & Burke, 2021). Over the last decade, efforts to strengthen the nation's workforce have led to government, private, and corporate funding

agencies initiating programs aimed at reimagining the STEM pipeline (Allen-Ramdial and Campbell, 2014). This is made possible through instituting programs that redress historical inequities, improve K–12 and postsecondary mathematics, science, and technology education, and increase participation of traditionally underrepresented groups in STEM fields (Greene, et al., 2006). Some of these programs have been claimed to be beneficial, although little is known about how they achieved their goals. Furthermore, there seems to be little compelling evidence of their cumulative impact in terms of improving the quantity or quality of the STEM pipeline (Greene, et al., 2006). Nonetheless, the development of solid workforce infrastructure plans that both ensure and extend access to high-quality STEM education is an economic priority for governments. Approaches and methodologies for evaluating STEM education programs that can make a difference will go a long way in setting the pace for the future of the field.

STEM education, which fuels the nation's economy, is described as teaching, learning, and related activities in the fields of science, technology, engineering, and mathematics in both formal and informal settings, from preschool to post-doctoral levels. (Gonzalez, & Kuenzi, 2012). Unquestionably, strong STEM education is essential for the United States to maintain its economic prosperity and global competitiveness (Breiner, Harkness, Johnson, & Koehler, 2012; Schwab, 2009). As a result, it's critical to understand the debate over evaluation approaches and methodologies, which serve as the foundation for STEM education programs fulfilling their goals (Ashby, 2006).

Roots of STEM Education Initiatives

Historically, STEM Education programs have roots in the educational agenda of the National Science Foundation (NSF), a federal funding agency that started investing in projects such as the Technology for All Americans Project (TfAAP), Content for the Study of

Technology (STL), and Advancing Excellence in Technological Literacy during the early 2000s (Dugger, 2010). NSF initiatives were largely driven by the political calculus of the U.S. Congress, which was propelled by professional and economic pressure to increase opportunities for learners who can participate in the STEM field (Blackley & Howell, 2015).

Since the 1990s, national and international conferences have focused on STEM education and its evaluation efforts (or lack thereof), with many types of discussions taking place within the realm of these gatherings. Today, STEM educators, scientists, administrators, policymakers, program managers, principal investigators, federal funding agencies, and other stakeholders continue to discuss STEM education-related topics. STEM practitioners participate in conversations aimed at increasing diversity and increasing minority participation, addressing gaps in STEM education, developing STEM literacy for everyday life, and promoting citizen empowerment for a more competitive workforce in the twenty-first century (Heil, Pearson, & Burger, 2013).

To accomplish its objectives, the U.S. government invests billions of dollars annually in federal funding towards revamping STEM education. However, the huge financial commitments have not yielded matching results (Anon, 2013). The lack of desired outcomes generated questions and a growing concern among stakeholders regarding federal resource allocation and efficient utilization to better serve STEM initiatives (Katzenmeyer & Lawrenz, 2006). The STEM educational challenge is further intensified by the complexity of federal efforts toward STEM enterprises and some government agencies lack coherent vision or careful oversight of goals and outcomes (Anon, 2013). Consequently, calls have been made for the evaluation of existing STEM education to better understand the current status of the field before organizing new federal programs and funding more projects that yield few benefits (Anon, 2013).

STEM Education Initiatives and Evaluation Discourse

Funding remains a top priority in discussions surrounding STEM educational program evaluation. The National Science Foundation (NSF), the Department of Health and Human Services (HHS), and the Department of Education (DOE) constitute the top three agencies leading federal spending on STEM education (Council, 2007). The Inclusion across the Nation of Communities of Learners of Underrepresented Discoverers in Engineering and Science (INCLUDES) program represents NSF's largest investment to transform STEM education and career pathways of citizens across the country. These three agencies (NSF, HHS, and DE) manage five of the STEM big efforts including (a) the Ruth L. Kirschstein National Research Service Awards; (b) the National Science and Mathematics Access to Retain Talent (SMART) grants; (c) the Mathematics and Science Partnership program; (d) Discovery Research K–12, and (e) the Graduate Research Fellowships program. Through its funding initiatives, NSF supports STEM pedagogy, engages in research capacity building, institutes fellowship programs, and promotes the goals of Broadening Participation at MSIs (Minority-Serving Institutions). In line with its goals of increasing diversity and Broadening Participation in STEM, NSF annual investment funding for MSI programs totals $136 million, including $46 million to the Louis Stokes Alliances for Minority Participation (LSMP), $35 million for the Historically Black Colleges and Universities (HBCU), $15 million for the Tribal Colleges and Universities (TCU), and $40 million for the Hispanic Serving Institutions (HSI) (Federal STEM Education Funding, 2020). Besides the labor implications and economic imperativeness of STEM enterprise, the political calculus advanced by the U.S. Congress on funding streams for STEM education programs suggests its priority lies in promoting equity and success for diverse learners. For example, out of an average of 200 STEM education programs operated by about a dozen

different agencies investing approximately $3 billion annually in STEM initiatives, it emerges that NSF votes $21 million for the INCLUDES, a program designed to mirror current diversity realities of U.S. population and tackle the incessant underrepresentation of women and minorities in STEM fields (Anon, 2013). Consequently, as the dialogue in STEM education programs gathers momentum, so does the scrutiny and accountability in the form of evaluation efforts.

Discourses of STEM Education at Evaluation Meetings

Discourses can be defined as a technique of forming or maintaining relationships within social and professional circles using language, thoughts, values, actions, and encounters that occur in the "right" locations and at the "right times" with the "right" people." (Gee, 2004, p. 34). While the term "discourse ", with a lowercase "d", exists in everyday use of language or communication exchanges like telling stories or having a dialogue or conversations with different people in each setting, uppercase "D" for Discourses goes beyond linguistics and symbolic operations of language (Gee, 2004).

The central element of Discourses is identification with a particular group or community that is immersed in a mixture of social, academic, and political institutions, where interactions are made possible through philosophical thought processes, multipurpose buildings fit for stimulating intellectual dialogues, theaters, classrooms, laboratories, and all forms of offline and online technology for engaging in conversations of common interests. Discourses are characterized by a set of belief systems, practices, or knowledge bases that construct and shape practitioners' views of reality and provide a shared way of understanding their world (McCloskey, 2008). Discourses with a capitalized "D" surface in an evaluator's identity or membership in a social network or professional organization and will be used in this study to

interrogate conceptions of approaches and methodologies within the STEM Education Program evaluation presentations at the American Evaluation Association (AEA).

To better understand how STEM education program evaluation works, it is crucial to know more about evaluation approaches and methodologies. Different types of meetings and events, including conferences, colloquia, congresses, plenaries or general sessions, seminars, symposiums, workshops, etc., are organized regularly throughout the year by academics, theorists, practitioners, program staff and managers, policymakers, and other stakeholders who may belong to various professional organizations within the field to present their work and discuss the current and future states while reflecting on the history of the evaluation field.

Among the dominant professional and sociopolitical communities, like the American Evaluation Association (AEA), American Educational Research Association (AERA), Center for Culturally Responsive Evaluation and Assessment (CREA), National Council on Measurement in Education (NCME), and the United States Congress, Discourses on STEM education and evaluation efforts are usually front and center. The familiar presence of the STEM enterprise as a remedy for the ongoing challenges in K–12 education, ample job opportunities and STEM career paths for STEM majors, a technologically literate citizenry, a thriving economy, and America's competitiveness in a global economy all explain the growing interest in STEM education (White, 2014). Evaluators in the field of STEM who attend workshops, seminars, webinars, symposiums, and conferences are thus tasked with engaging in a critical dialogue that demonstrates a thorough understanding of their practice and the ability to effectively communicate it through effective evaluation approaches and methodologies.

Discourses of Evaluation Approaches in STEM Education Programs

A plethora of approaches exists in the evaluation literature, and a definition or conceptualization of the evaluation approach Discourse depends on the evaluation questions or challenges facing a given project or the issues at stake among the evaluands and its program community. The Discourse surrounding evaluation approaches in STEM education is broad and integrative, frequently drawing from multiple and intersecting contexts and settings. For example, the systems thinking approach has been developed to address questions about interconnections among salient factors in complex school systems and outside the partnership in which K–12 STEM education, higher education universities, informal and community education organizations, and businesses collaborate on a project (Saxton et al., 2014).

Based on a preliminary review of the literature, the types of evaluation approaches currently being utilized are shaping and guiding the scope and type of Discourses surrounding STEM education evaluation. Evaluators working in the STEM fields are engaging in dialogue and seeking answers to questions based on three main components of their STEM education initiative: (a) program context (b) the implementation process; and (c) program outcome (Kellogg, 2004). In addition, evaluation approaches streamline the types of questions asked in program contexts in terms of program capacity building, and relationship constituents such as knowledge and power dynamics between evaluators, program managers, staff, program participants, other stakeholders, and funders. During the implementation process, questions related to how a STEM education program works can reveal the intrinsic values and assumptions that are rooted in the evaluation as well as the project itself. Does the evaluation approach seek to clarify questions about: a) which questions get asked? b) Who gets asked the questions? and c) what order the questions are asked? The program outcome regarding what the STEM program is

set to achieve and whose interests it aims to serve can benefit from the evaluation approach adopted.

In addition, STEM evaluation scholars recognize the field goes beyond the STEM ecosystem and education domains. The STEM enterprise has the potential to reach people from diverse backgrounds, thus making it possible for evaluators to give voices to evaluation approaches that align with evaluation questions while centering the value stances and educational needs of program contexts (Boyce, 2017; Greene, Boyce, & Ahn, 2011; Greene, et al., 2006; Hall, Freeman, & Colomer, 2020; MacDonald & Bourke, 2017). To understand the major evaluation approaches used in STEM education, it is important to highlight their specific contexts. For example, several authors have utilized the concept of culturally responsive evaluation approaches in programs designed to broaden the participation of underserved populations and underrepresented minorities in STEM education (Johnson, 2005; Garibay & Teasdale, 2019; Mertens & Hopson, 2006; Sayler, Apaza, & Austin, 2005).

Overall, their discussion showed that implementing a culturally responsive evaluation approach to STEM education while being attentive to program culture, program participants, and key program stakeholders works well toward achieving the diversity and equity goals of STEM education programs. However, cultural, and contextual factors are never fully accounted for, and their roles are not well understood by culturally responsive evaluators working in the STEM field. This may result in inadequate exploration of existing belief systems that impact indicators of success, validity, or bias and create gaps in knowledge where data collection, analysis, and interpretation are concerned (Johnson, 2005).

Furthermore, there is a cultural Discourse in the STEM field that numbers and statistics “don’t lie” and there is often a preference in academia for numerical data, quantitative methods,

and “objectivity.” A lot of our quantitative data analysis methods are built around concepts like a “mean” and a “standard deviation,” but people’s experiences and outcomes can vary in important ways, so the idea of running statistics using a grand mean or any other standardized mathematical unit operation has severe limitations. Nonetheless, practitioners in the STEM field continue to engage in this practice as a result of its widespread professional acceptance and institutionalization.

STEM evaluation scholars have also proposed a values-engaged educative evaluation approach that goes beyond just the prescription of arrays of strategies, but one that centralizes the values of the program and prioritizes stakeholder engagement (Greene, Boyce & Ahn, 2011). What is distinctive about the values-engaged, educative evaluation process is not the set of methods or approaches to evaluation design, but rather its value commitments and engagements. Accordingly, values-engaged, educative evaluators are encouraged to use and select from the full repertoire of designs and methods available that are relevant to the evaluation questions at hand and to the demands and variables of the given program context.

Discourses of Evaluation Methodologies in STEM Education Programs

The current Discourse surrounding evaluation methodologies in STEM education emerges from the theoretical debates between qualitative and quantitative methodologies (Lawrentz & Huffman, 2006). On the one hand, qualitative methodologies are based on the idea that there are numerous perspectives on reality, depending on the researcher’s perspective (Pascale, 2011), and that subjective rather than objective worldviews are preferable (Creswell & Poth, 2016). On the other hand, quantitative methodologies are based on the worldview that objective reality exists independent of human perception (Sale et al., 2002), where the researcher

can study a phenomenon without influencing it or being influenced by it (Denzin & Lincoln, 1994).

The discussion in the STEM evaluation literature highlights methodological pluralism in evaluation practices and shows the strengths and weaknesses of different methods, but the debate has paid little attention to the philosophical underpinnings of methodologies that inform and shape the choice of methods. Scholars in support of scientific and quantitative methods have argued for the appropriateness of rigor, citing sample sizes, the statistical fit of the data, meanings, and the interpretation of results and findings as some of the rationale for valuing quantitative approaches (Thomas & Harden, 2008). On the other hand, staunch defenders of qualitative research have contrasted validity in quantitative research (internal validity, external validity, reliability, and objectivity) with evaluation's trustworthiness (credibility, transferability, dependability, and confirmability) and acknowledged that they are different, but qualitative methods are no less rigorous when compared with quantitative methods (Lincoln & Guba, 1985).

Despite the dominant debates rolling from quantitative methods, which involve measuring and analyzing numerical data based on statistical tools, to qualitative methods that utilize descriptive and interpretive instruments, a mixed-methods evaluation that utilizes a combination of quantitative and qualitative philosophies is gaining traction in the STEM field. Evaluation scholars in favor of the third approach believe that multiple or mixed-method designs are better suited to tackle inequities, diversity, cultural, and numerous other challenges facing underserved and underrepresented minorities in STEM (Mertens & Hopson, 2006) while considering the multidisciplinary and integrative STEM field (Lawrenz & Huffman, 2006).

Mixed methods also strengthen evaluation in two ways: (a) through component designs that enable data triangulation and complementarity, and (b) through an integrative design that is

iterative, and transformative (Caracelli & Greene, 1997). While the discussions about the choice of evaluation methods are active and distinguishable for the most part, the philosophical assumptions undergirding them remain unclear. A preliminary review of the evaluation literature suggests the language of evaluation methodologies used in STEM education is framed in multiple perspectives (Feuer, Towne, & Shavelson, 2002; Gullickson & Hanssen, 2006). In the belief and value systems of the evaluator (Greene, et al., 2006; Malyn-Smith, 2014), in using quotations to strengthen arguments (Mertens & Hopson, 2006), and in surfacing institutional and researcher biases in practice (Committee on Equal Opportunities in Science and Engineering [CEOSE], 2004; NSF, 2002; U.S. Department of Education, 2004). Consequently, it is imperative to draw a distinction between methods and the focus of this study, methodologies, and to expound on the Discourse surrounding STEM education evaluation methodologies.

Problem Statement of the Research

The STEM fields are critical to the United States' ability to create job opportunities and sustain economic prosperity. With projected growth rates of 13% between 2012 and 2022, STEM occupations are expected to outpace other professions (Fayer, Lacey, & Watson, 2017). The National Science Foundation (NSF) has launched numerous initiatives over the past 30 years to develop and strengthen STEM education, with a focus on Broadening Participation to increase participation from underrepresented groups and diverse institutions (NSF, 2020).

Despite significant investment in STEM education programs, there remains a limited understanding of the Discourses surrounding evaluation approaches and methodologies utilized by STEM evaluation practitioners. Burbules (2004) posits that constructive Discourse is necessary to optimize evaluation, highlighting the need for a deeper level of conversation among

STEM education professionals. Professional meetings such as the American Evaluation Association (AEA) provide opportunities for this conversation to take place.

In conclusion, STEM fields have the potential to create job opportunities, sustain economic prosperity, and deliver leadership in scientific excellence and technological advancements. The NSF has launched several initiatives to develop and strengthen STEM education, with a focus on increasing participation from underrepresented groups and diverse institutions. However, there is a need for constructive Discourse and optimization of evaluation approaches and methodologies utilized by STEM evaluation practitioners to ensure the effectiveness of STEM education programs.

Purpose of the Study

This study examines the Discourse of STEM education evaluation from the perspective of practitioners. It investigates the prevailing evaluation approaches, including values-engaged, educative, culturally responsive, and utilization-focused, and the various methodologies used, such as quantitative, qualitative, and mixed-methodologies, case study methodologies, survey methodologies, focus groups, and interview methodologies. The study identifies gaps in the approaches and methodologies used by evaluation practitioners, shedding light on the types of evaluation approaches and methodologies discussed theoretically and utilized practically in the field of evaluation.

To better understand the Discourse around STEM education evaluation, the study reviewed and analyzed abstracts from the AEA conference STEM TIG and conducted individual interviews with STEM education evaluators. The interviews provided insight into practitioners' impressions and experiences with their work, illuminating perspectives on the prevailing

Discourse in STEM education evaluation and the field's future direction amid the evolving Covid-19 epidemic.

The study's findings contribute to the discussion surrounding the current state and future direction of STEM education program evaluation and add to the discussion of approaches and methodologies utilized by STEM education evaluation practitioners. The study is significant because there is a lack of knowledge in the evaluation literature regarding the Discourses surrounding approaches and methodologies utilized in the STEM education evaluation. The research questions that guided this study include the following:

1) What was the Discourse of STEM Education Evaluation presentations at the annual American Evaluation Association conference from 2014 to 2019?
    a) What are the main evaluation approaches and methodologies utilized in these presentations?
    b) What are the differences between evaluators who work in academic institutions versus practitioners in private and public organizations?
2) What are STEM evaluators' thoughts and impressions about the current and future state of the field?
    a) To what extent do the initial findings from the STEM evaluation literature of Discourses resonate with STEM education evaluators about the current and future state of the field?

Significance of the Study

The devastating Covid-19 pandemic has put the world's population and STEM education as a field amid a significant health and humanitarian problem. This ongoing state of affairs when layered on existing challenges in areas of homeland security (Adamski, Kline, & Tyrrell, 2006) and national security (Sindico, 2007), climate and energy crisis (Karl, Melillo, & Peterson, 2009;

Melillo, Richmond, & Yoke, 2014; Munich Re, 2015), health and environmental protection issues (Elsea & Mason, 2008; Flynn, 2011; NCA, 2014), economic growth (O'Sullivan & Ramsay, 2015), and food security (Bailey, Benton, & Challinor, 2015), send strong signals to the evaluation field to shed more light on the current Discourses and reimagine the future of the STEM enterprise. As these pressing challenges become direr, the need for skilled workers increases, pressure mounts on the labor force for more competent professionals (National Academies Press, 2010), calls for developing the STEM enterprise grows, and global competition for expanding the talent pool in STEM education becomes more intense (Davis & Hart, 2010; Shachar 2006).

Previously, STEM occupations projected growth of 17% from 2008 to 2018, compared to 9.8% growth for non-STEM occupations (Langdon et al. 2011). Conversely, STEM education has a history of low participation of women and people of color, and there is an urgent need for personnel and progress in science and related fields (Greene et al, 2006). Several authors (Boyce, 2017; Boyce & Rivera, in press; Greene, et al., 2011; Greene, et al., 2006; Hall, et al., 2012; Reid, 2020) have proposed the values-engaged, educative (VEE) evaluation approach to address the needs of STEM education. Despite the potential of the VEE and similar well-intentioned approaches, there exists a chasm between theoretical and empirical literature. Additionally, there is a gamut of methodologies available to evaluators for consideration when designing and implementing an evaluation. Furthermore, this study interviewed hard-to-find STEM education evaluators who are knowledgeable in the field, well-trained, have become specialists, or have emerging voices within the STEM education evaluation community. Given the inadequacy of highly skilled evaluators involved in STEM education programs and initiatives (Katzenmeyer & Lawrenz, 2006), the perspectives of the interviewees make it more important than ever in

conversations about the evaluation approaches and methodologies used in STEM education programs. In other words, the perspectives of STEM education evaluators interviewed simply optimize the current and future states of the STEM education evaluation field.

This investigation reviewed and critiqued the current STEM education evaluation Discourse within the field. Concretely, this work explored (1) types of evaluation approaches and methodologies discussed theoretically and utilized empirically in the literature, and (2) the types of evaluation approaches and methodologies evaluators are using in their practice. Given the ubiquitous influence of science and technology and allied subjects, it was imperative to have illuminated the Discourses of STEM education evaluation by foregrounding methodological approaches and a variety of methodologies utilized by practitioners in the interest of STEM education that continues to expand at an exponential rate.

CHAPTER II: LITERATURE REVIEW

STEM Education Evaluation Literature

The Discourse of Science, Technology, Engineering, and Mathematics (STEM) education regarding evaluation approaches and methodologies continues to raise questions in recent times in the evaluation literature (Allen & Peterman, 2019; Alemdar, Cappelli, Criswell, & Rushton, 2018; Boyce, 2017; Garibay, & Teasdale, 2019; Greene, et al., 2006; Havice, Havice, Waugaman, & Walker, 2018; Huffman, Lawrenz, & Lawrenz, 2006; Mertens & Hopson, 2006; Moon, Utschig, Todd, & Bozzorg, 2011; Saxton et al., 2014). This could be ascribed to the ubiquitous influence of STEM enterprises on the nation's educational needs, the creation of a skilled labor force, gainful employment in a diverse economy, and opportunities for more tech-savvy workers for the growing hi-tech industries. In meetings and professional gatherings across the country, STEM education program evaluators contribute to the discussion on U.S. education, funding, global economic competitiveness, strengthening of national security, immigration policies, harnessing of innovation, trade, and Broadening Participation of underrepresented populations to increase their talent pool in STEM fields for a more prosperous America.

While the perceived roles of administrators, policymakers, and parents have received wide attention, it remains unclear within the STEM Discourse what set of evaluation approaches undergirds evaluators' philosophical decision-making process in support of STEM education programs. Moreover, it is becoming increasingly important in the evaluation of STEM education programs amid the Covid-19 pandemic for evaluators to reimagine the array of evaluation methodologies deployed in practice. In the STEM evaluation literature and other fora such as the STEM TIG at AEA, evidence of attention to methods abounds in the discussion, dialogue, and engagement among STEM evaluators. However, as Carter and Little (2007) suggested

methodologies that "justify, explain, and help us understand" the different ways of thinking that shape evaluators' choice of research objectives, research questions, and study design remains elusive in the STEM evaluation Discourse and need further investigation.

Evaluators and other professionals working in STEM education enter the field from diverse backgrounds, work in varying contexts and settings, subscribe to multiple worldviews, and make use of information sources and strategies they believe are reflective of a given program's overall culture. Evaluation practitioners may value quantitative, qualitative, or mixed methods approaches which show up in their thought process in evaluation objectives, questions raised, study design, implementation, report writing and dissemination, and shape communication within the Discourse of STEM Education Evaluation.

In conducting a literature review for this dissertation, I started by overviewing the concept of STEM education and provided an operational definition for its meaning. Thereafter, I discussed some of the dominant STEM education initiatives designed by the U.S. federal government and other organizations to equip citizens with STEM literate skills, sustain the local workforce, and strengthen U.S. leadership and global economic competitiveness through attention to Broadening Participation programs (Dasgupta & Stout, 2014; Estrada, Eroy-Reveles, & Matsui, 2018; Fealing, Lai, & Myers Jr, 2015; Holloman et al., 2018; Hrabowski III, 2012).

This was followed by examining the Discourse surrounding how STEM education programs operate within its social, economic, and political ecosystem and project settings, and paying attention to the core attributes utilized in determining a quality STEM education program (Yarbrough, Shulha, Hopson, & Caruthers, 2010). This study briefly discussed funding issues in STEM education, and the role played by federal agencies such as the NSF in pioneering programs and supporting grants. This was necessary because funding remains directly tied to

availability and opportunities to access quality STEM education programs which have seemingly been out of step with the larger community it desires to serve, therefore, creating evaluator awareness of it as a major Discourse increases the incentive to talk about in meetings and gatherings (Tanenbaum, 2016).

In summary, this chapter provides a broad overview of the existing literature related to my research idea. I identified themes, trends, and tensions in the approaches, methodologies that undergird the design, and findings of the Discourses surrounding STEM education evaluation. I reviewed and critiqued the existing STEM education evaluation literature and interrogated the contributions of the scholars related to my topic. I highlighted the methodological strengths and weaknesses of the research studies within the STEM education evaluation literature. Specifically, in this chapter, I reviewed and critiqued the STEM education literature concerning Discourses of evaluation approaches and methodologies utilized.

As Congress, funding agencies, scholars, practitioners, and others in the STEM community continue to demand evaluation of programs, evaluation approaches and methodologies used in the STEM education assessment will continue to gain attention and become more prominent. My analysis generated conversations and stimulated engagements that seek a better understanding of the types of approaches and methodologies that are prevalent within the context of STEM education evaluation. In the process, I highlighted approaches and methodologies that are discussed theoretically versus empirically within the STEM education evaluation literature, thus producing a clearer picture of the current state and future of the STEM education evaluation field.

The Conception of STEM Education

The term "STEM education" has been used in discussions (Allen & Peterman, 2019; Garibay & Teasdale, 2019; MacDonald, Huser, Sikder, & Danaia, 2020), in multiple languages **(**Compton, Baizerman, Preskill, Rieker, & Miner, 2001; Han, 2015; Reis, Solovey, Henner, Johnson, & Hoffmeister, 2015), at conferences (Huffman, Lawrenz, Thomas, & Clarkson, 2006; Newcomer, 2004;), during engagements (Franco & Patel, 2017; Kim et al.,2015), when interactions are taking place (Greene, et al., 2006; Rahman, 2020; Thompson et al., 2016), and in general, when communicating (Allen & Peterman, 2019; Hoeg & Bencze, 2017; Katzenmeyer & Lawrenz, 2006; Reeves, Bobrownicki, Bauer, & Graham, 2020) with others in different settings by educators, scholars, evaluation practitioners, project managers, policymakers, and numerous stakeholders interested in the STEM enterprise. Many stakeholders use STEM education as a currency to think through, dialogue with, and shape understanding of past and present evolving theories, as well as inform future Discourses related to STEM education practices.

The literature investigated in this study was the interdisciplinary or collaborative approaches to Discourses in the evaluation of STEM education programs as a visible enterprise through the STEM pipeline, including informal STEM settings (Barakos, Lujan, & Strang, 2012; Brown, Brown, Reardon, & Merrill, 2011). I examined the four disciplines of science, technology, engineering, and mathematics as an integrated field and fused the theories, practices, teaching, and learning experiences of all stakeholders in the STEM enterprise. A working definition that referenced STEM education as the teaching and learning in the STEM field, including educational activities across all grade levels, from pre-school to post-doctorate, in both formal (e.g., classrooms) and informal (e.g., after school programs) settings were highlighted throughout this review (Gonzalez & Kuenzi, 2012). This literature review paid attention to the

Discourses of STEM education evaluation driving the debate in formal and informal settings on wide-ranging issues such as scholarship opportunities for attracting and retaining STEM majors towards a career path in the field, and the provision of funding for research and evaluation across the STEM education spectrum (Act, 2010). The review considered the concept of STEM education evaluation Discourses in professional meetings where evaluation theorists and practitioners gather to have conversations with their colleagues and peers on topics that advance knowledge and make use of evaluation and research (American Evaluation Association [AEA], n.d.).

STEM education begins in Pre-Kindergarten, where parents start to read to their two-year-old children and help them learn simple skills with the goal of early childhood development. At primary school, childhood development and educational progress begin to take shape, with parents, and teachers supporting children in terms of cultural resources, helping with school homework, and monitoring academic performance starting in 4th-grade STEM subjects like math and science. By the time students reach middle school, conversations around childhood development, educational progress, and finances become front and center. Stakeholders with interests in education start conversations and want answers to questions about U.S. 8th graders' proficiency in math and science within and across each state of the nation. At high school, there is much emphasis on how U.S. 12th-grade students score in math and science, how proficient they are in math and science, what highest-level math and science courses they are permitted to take in 9th grade, the identity of students who take advanced placement courses in math and science, and the percentage of U.S. public high school students who graduate on time and head to college. College-level Discourses about STEM education take multiple forms. Conversations center around college entry enrollment, several degree options available for students in STEM

fields, STEM major retention rates, Broadening Participation of underrepresented minority students, the inclusion of women and students with disabilities, graduation rates, pursuing advanced degrees in STEM, or following pathways to STEM career, financial aid spending for undergraduate students and funding from states at public research universities.

In the last ten years, STEM jobs available in the marketplace have exceeded the projected employment expectations of 18.7% compared to 14.3% for all other occupations (U.S. Bureau of Labor Statistics, 2018). Students who graduate with STEM majors and those with practical and requisite skills associated with STEM disciplines are well-positioned to fill the growing demand of the STEM workforce. Consequently, as a key driver for domestic STEM labor supply, and coupled with its potential to equip citizens with knowledge and expertise to compete in the global economy, the critical role of a STEM education system cannot be overemphasized, and the time is ripe to shine the light on Discourses surrounding evaluation approaches and methodologies of the STEM efforts.

STEM Education Programs

Programs in STEM education are primarily designed to promote learner achievement, increase equitable access, and bring about an improvement in the educational and overall life experiences of American citizens. These programs are tailored to be responsive to citizens' everyday needs across America and all over the world (Epstein, 2006; Granovskiy, 2018; Tushnet, 2003; Gonzalez, 2012; Council, 2007). Education programs in STEM are anchored on the federal government's "Strategic Plan" for success and call to action that was established in 2018 to provide all citizens the opportunity of access to quality education in STEM while fostering American STEM literacy, and leadership in a competitive global economy. STEM education programs feature early childhoods to graduate schools and serve to equip K–12

learners with basic STEM coursework and skill set, provide knowledge, research, scholarship, and grant opportunities for undergraduates, and make a pathway for them to graduate school or lead to a career in the STEM field.

The federal government, through the Committee on STEM Education (CoSTEM), prioritizes the evaluation of STEM education programs for effectiveness (MacIsaac, 2019). The government relies on STEM education programs for development through collaborations across the STEM ecosystems. Many of the STEM education programs are situated in settings and contexts such as PreK–12 Education, informal education (e.g., youth development programs), community colleges, and university faculty (such as HBCUs, MSIs, and HSIs) (Committee on STEM Education, 2018). A sample of the programs has been summarized below (Table 1). The information was arranged according to the program's name, the funding source, a real-world program example, and setting.

Funding STEM Education Programs

Funding for STEM education programs is essentially driven by U.S. government investments in research and development (R&D), education, and training (E&T) activities (Anon, 2013; Gonzalez, 2012; Katzenmeyer & Lawrenz, 2006). The first step in understanding the effort to evaluate the capital-intensive STEM initiatives is to establish a common definition for the purpose and scope of the two groups responsible for overseeing the federal STEM education research portfolio. The National Science and Technology Council’s (NSTC) Education and Workforce Development Subcommittee (EWD) have defined the boundaries of activities under them, these include: (a) determining the current and future needs of the STEM ecosystem that can sustain the federal workforce, (b) identify and support federal programs to contribute to

the national STEM workforce, and (c) propose solutions, actions, and initiatives that address those needs (Berch et al., 2006).

The early 2000s witnessed a push to address the gap experienced by U.S. students' achievements in the STEM disciplines when compared to other advanced countries. A response to that gap gave life to the STEM acronym, and states in the nation began to acquire public and private grant funding to support STEM education. As the drive to join the STEM race gathered momentum, more states began to secure funding for K–12 education, and the No Child Left Behind (NCLB) Act was passed in 2001 (Malyn-Smith, 2014). Concomitantly, the demand for funding for STEM education across grade levels, disciplines, and learning environments, ranging from K–12 schools to summer camps and after-school programs to postsecondary and research institutes, and informal education like the libraries and museums began to ramp up. The federal government together with its funding agencies like the National Science Foundation (NSF), the National Institute of Health (NIH), the U.S. Department of Education (DE) and the National Aeronautics and Space Administration (NASA) led the way in providing research and evaluation grants for various formal and informal initiatives.

The NSF plays a significant role in funding STEM education programs as the only federal agency with a primary focus on supporting education across the integrated STEM field. The link between STEM education systems and prosperous workforce as key economic drivers in the nation is well documented. Just a little over ten years ago, a report by the National Science and Technology Council (NTSC) showed a $3.4 billion investment in an average of 179 STEM education programs across approximately 14 federal government agencies (Gonzalez, 2012). This report reveals three agencies, NSF, ED, and the Department of Health and Human Services (HHS) received the biggest funding.

Table 1. Typical STEM Education Programs

STEM Program	Funder/ Agency/ Partners	Example
Fostering Partnerships to Ensure Broader Participation in STEM	NSF INCLUDES	A partnership between a State University and School District to design K–12 computer science curriculum
Connecting the People and Tools for Transdisciplinary Learning	NASA Fellowship Programs engage HBCUs, HSIs, and K–12 schools in STEM	Louisiana Space Grant manages High-Altitude Student Platform (HASP).
Integrated Approaches Computational Thinking Instructions	NSF supports teaching modules and courses that infuse computational thinking	A diverse range of students and teachers benefits from computational thinking skills in STEM
Teaching Data Science through Community Science	Multisite STEM education programs with local stakeholders and Federal partners	Multisite community science projects using GIS mapping technologies
Documenting the Participation of URMs STEM Programs	Government ensures all STEM programs are inclusive and welcoming to diverse groups	AEOP utilizes a shared definition of “underserved” across its program portfolio for reporting consistency

Furthermore, the NSF is the largest agency in terms of funding and has the most STEM programs. Generally, funding for R&D and E&T activities fluctuates, but R&D shows more upward trends over time. However, scholars and practitioners alike have called for more investment in STEM education in general, and within the STEM education "pipeline" (pre-kindergarten to post-graduate education) exactly. This call does not only show up in the policy rationale that drives funding but has important implications for the Discourse of STEM Education Evaluation.

As funding trends at federal agencies continue to raise pertinent questions for lawmakers, the U.S. government intensifies efforts aimed at the coordination of more than 83% of the overlapping STEM education programs (Anon, 2013). The federal government identifies which STEM education programs to consolidate effort on and other ones to eliminate from its portfolio to increase the efficient use of finite resources. The role of evaluation in this process is becoming increasingly critical and the demand for clarity and specificity in terms of evaluation approaches and methodologies used in practice is gaining more attention. Moreover, having the government subscribe to rigorous evaluation of STEM education programs will provide the much-needed understanding and clarity in terms of what works and what does not work. In addition, evaluation efforts will be able to address gaps in research agendas, assist in decision-making, and prioritize efficient utilization of federal funding support for the present and future investments in basic and applied STEM education initiatives.

Consequently, debates and discussions continue to attract attention at evaluation meetings about the specific roles of funding agencies within the federal STEM education portfolio. Additionally, congressional hearings are underway deliberating on the impact of budget cuts on research and evaluation activities. Another major concern is what the future of STEM education

programs and evaluation may look like as the federal government contemplates a change in funding policy directions. Program managers or directors, researchers, principal investigators, educators, scholars, and other stakeholders responsible for developing and implementing STEM education programs continue to play a pivotal role in working with evaluators who seek to explicate and contextualize Discourses in STEM education evaluation. Evaluators' contribution to the local, state, and national dialogue on funding STEM education and evaluation initiatives remain invaluable in advancing the STEM field.

STEM Education Evaluation Literature—Approaches and Methodologies

Studies for this portion of the literature review were based on the following inclusion/exclusion criteria: (a) peer-reviewed articles; (b) publications between 1994-2021; (c) studies conducted within the integrated STEM field and STEM evaluation literature; and (d) available through online databases. Precisely, the following criteria were considered for the inclusion of papers or articles in the review:

1. Is the paper or article peer-reviewed?
2. Do the program evaluation approach, or methodologies include Discourses on the STEM disciplines?
3. If so, does the paper contain elements of science and mathematics Discourse?
4. Does the article provide empirical evidence or sufficient theoretical background regarding the evaluation approaches or methodologies described?
5. Do the authors present sufficient information or insights that are likely to contribute to the Discourse of STEM Education Evaluation?

The articles did not have to meet all five criteria for inclusion in the review process but meeting the first criteria is necessary. Meeting additional criteria increased their chance of final inclusion.

The Analysis Process

The following steps describe the basic elements of how I analyzed the data for the literature review. My initial search of the broader review of the STEM evaluation literature to determine the extent of reporting on frequently utilized evaluation approaches and methodologies within the STEM context yielded over 500 reference citations. They were then screened using unique keywords in isolation and in combination to search online databases such as Academic Search Complete (EBSCO), ProQuest Education, ERIC, and Google Scholar. A good number of the citations were from the grey literature or publicly available evaluation studies or funded project reports through NSF or National Aeronautics and Space Administration (NASA), but a few of them were from published peer-reviewed journals. Using a range of terms including various combinations of the following: "Evaluation of STEM educational programs" "STEM evaluation programs," "STEM evaluation," "Evaluation approaches and STEM," "Evaluation of STEM," "Science and engineering evaluation approaches," "Science evaluation approaches" and "Science and Math," "Evaluation methodologies and STEM," "Science and engineering evaluation methodologies," "Science evaluation methodologies" and "Science and Math," "STEM education and Discourse of evaluation approaches," "STEM education and Discourse of evaluation methodologies" and "STEM education programs and Discourse of evaluation approaches and methodologies."

Following this initial review, full-text copies of 100 of these citations were obtained to assess their potential for inclusion in the last review. After further analysis, a total of 48 sources including 34 articles from eight journals, 4 books or book chapters, and 10 conference proceedings or reports was retained in the final analysis. The remaining 52

citations were excluded because they did not meet the inclusion/exclusion criteria outlined above.

The breakdown of the articles includes twelve reference articles/reports in the New Direction for Evaluation (NDE) journal; 6 articles/reports in the American Journal of Evaluation (AJE); 1 article/report each in the Canadian Journal of Program Evaluation (CJPE), Journal of Multidisciplinary Evaluation (JME), Evaluation Review (ER), and The Evaluation Exchange (EE); 7 articles/report in Journal of Evaluation (E); and 5 articles/report in Advances in Program Evaluation (APE). The papers reviewed date back to 2003-2021. With the exceptions of 2006 and 2011 with multiple papers each, only a handful of papers have been published in the 20 years inclusive.

Each article or report was reviewed multiple times to gain a better understanding of the context within which the STEM education evaluation Discourse was carried out. The types of debates and conversations going on in the field were instrumental to grasping the theoretical and empirical underpinnings of approaches and methodologies utilized in STEM education evaluation. Rather than use preconceived themes, articles and reports were read through to find recurring ideas or conceptions that speak to the Discourse of evaluation approaches and methodologies within STEM education evaluation.

There is an ongoing debate in the United States education sector concerning the best-fit approaches for conducting research and evaluation of STEM educational programs. For example, the U.S. Department of Education (2004) has prioritized science-based evidence evaluation methods and championed a “gold standard” of randomized controlled experimentation (Lawrenz & Huffman, 2006). Quasi-experimental designs with carefully matched comparison conditions, longitudinal designs, and regression discontinuity designs were also prioritized (Lawrenz &

Huffman, 2006). However, the STEM education evaluation literature suggests a different picture when it comes to the evaluation of STEM educational programs. Reports gleaned from the STEM education evaluation literature showed that restricting evaluation to "scientific" approaches is limiting (Maxwell, 2004), that advocating a gold standard for the evaluation of education programs has the potential for fallacy, and that assuming only the regularity view of causation is untenable. Conversely, evaluation scholars and researchers have argued for understanding and appreciation of multiple perspectives instead of promoting specific approaches (Feuer, Towne, & Shavelson, 2002). In my dissertation, I highlighted the Discourses of evaluation approaches utilized in STEM education programs and various methodologies that justify the choice of methods for alignment with components of an evaluation. The dominant evaluation approaches in practice have previously been identified and categorized in the literature and my dissertation will follow the framework that is consistent with the work of Stufflebeam and Coryn (2014).

Evaluation Approaches and Methodologies in STEM Education Contexts

When conducting this literature review, I identified four distinct evaluation approaches and three approaches that used a combination of two or more strategies. The driving force behind methodologies was largely powered by the evaluation questions being asked. The major approaches identified in the STEM education evaluation literature are (a) Values-Engaged, Educative (VEE) approach; (b) Citation Analysis (CA), and (c) Systems thinking/Systems-oriented. Three studies used a combination of two or more approaches including (a) Culturally Responsive Evaluation (CRE) and Systems-Oriented Evaluation (SOE); (b) Social Agenda (SA), Responsive (R), Constructivist (C), and Deliberative-Democratic (D-D); and (c)

Participatory/Collaborative (P-C) approaches to STEM education evaluation. One study used the "Evaluation Capacity Building (ECB)" (**Error! Reference source not found.**).

Table 2. Summary of Evaluation Approaches and Methodologies in the STEM Education Contexts

Study	Approach	Context/Focus	Methodology/Method
Hall, J.N., Ahn, J., & Greene, J.C. (2012)	Values-Engaged	Committed to Descriptive and Prescriptive Values	Case Study and Critical Reflections
Mertens and Hopson, (2006)	Social Agenda/Advocacy, Responsive, Constructivist, etc.	NSF STEM Projects Focused on Diversity and Cultural Issues	Case Study (Reflective analysis)
Lawrenz and Huffman, (2004)	Participatory/Collaborative Approaches to STEM evaluation	Two settings: K–12 Schools and Institutions of Higher	Surveys (web-based and paper and pencil), Observation
Boyce, A. S., (2017)	Values-Engaged, Educative Approach	Lessons Learned in Implementation of STEM Education Program Evaluations	Case Study Design: Prioritize Culture, Diversity, and Equity

Approaches and Methodologies Used in the Evaluation Literature

Evaluation Approaches

Some of the most frequently used and often cited approaches encountered in the evaluation literature include the following:

Decision/Accountability-Oriented Studies

This approach is used frequently in a forward-looking way to enhance a program and for reflection to assess its validity and worth. The approach's philosophical foundations include an objectivist commitment toward finding the best-fit solutions to context-limited problems and embracing ideals of a well-functioning democratic society, including those that promote equity, diversity, and inclusion, as well as fairness, social justice, and accountability (Stufflebeam, & Coryn, 2014)

Case Study Evaluations

Is a detailed description, analysis, and synthesis of a given program, project, or other initiatives? Evaluators using case study evaluation do not influence the programs; however, they examine programs in their geographic, cultural, organizational, and historical settings. Evaluators closely monitor internal operations and how inputs and processes are used to achieve program outcomes. This approach is used to investigate a wide variety of expected and unexpected outcomes or consequences of a program. It is used to examine a program at various stages as well as the overall program. It describes both central prevailing themes as well as deviations and outliers (Stufflebeam, & Coryn, 2014).

Utilization-Focused Evaluation

The utilization-focused approach is designed and tailored to ensure that program evaluation has a positive impact (Patton, 2013). It follows a process for making decisions

regarding an evaluation study in partnership or collaboration with a targeted population or group of priority users chosen from a larger collection of stakeholders to focus successfully on the original evaluation's intended uses (Stufflebeam, & Coryn, 2014).

Client-Centered Studies (or Responsive Evaluation)

This approach encourages evaluators to work with and for the support of a diverse group or community of participants including, for example, students, teachers, administrators, program developers and managers, and funders or funding agencies. They constitute the programming community in the sense that they support, develop, administer, or directly operate the programs under study and seek or need evaluators' counsel and advice in understanding, judging, and improving programs (Stufflebeam, & Coryn, 2014).

Evaluation Methodologies

Some of the most frequently used and often cited methodologies encountered in the evaluation literature include the following:

Quantitative Designs

STEM education quantitative evaluation designs with a positivistic philosophy, including utilitarian and nomothetic orientations that attempt to establish or indicate causality, are frequently used in research. However, due to the difficulties in applying random processes, many of the designs tend to be quasi-experimental rather than true randomized control experiments and thus have limitations (Stufflebeam, & Coryn, 2014).

Quasi-Experimental Design

Like an actual experiment, Quasi-Experimental Design seeks to establish a cause-and-effect relationship between an independent and dependent variable. A quasi-experiment, unlike a true experiment, does not rely on random assignment. Subjects are instead assigned to groups

based on non-random criteria. In circumstances where genuine trials cannot be performed for ethical or practical reasons, a quasi-experimental design can be effective (Stufflebeam, & Coryn, 2014).

Correlational Designs

Designs with Correlations are longitudinal and use historical data to forecast future patterns. Deviation from expected patterns is thought to demonstrate the effectiveness of any intervention. This type of strategy necessitates extensive databases that can track individual participants over time and where impacts other than an intervention can be considered random. These requirements are difficult to achieve, and designs are frequently compromised in one way or another (Stufflebeam, & Coryn, 2014).

Qualitative Methods

The philosophy behind qualitative approaches is one of description and emergent interpretation, with idiographic and intuitionist viewpoints. The possible answer or hypothesis is unknown ahead of time. In local STEM evaluations, qualitative methodologies are often utilized to collect in-depth and continuous perceptions of participants and stakeholders. Because it is the local example that is important, these evaluations transcend the typical complaint of lack of generalizability (Lawrentz & Huffman, 2006)

Case Study Designs

Are used to gain concrete, contextual, in-depth knowledge about a specific real-world subject. It allows researchers to explore the key characteristics, meanings, nuances, particularities, and implications of a case. The case study design, according to Yin (1994), must include five components: (a) the research question(s), (b) its propositions, (c) its unit(s) of analysis, (d) a determination of how the data is linked to the propositions, and (e) criteria for

interpreting the findings. Furthermore, according to Stake (1995), the number and type of case studies required in an investigation depends on the purpose of the inquiry. For example, an instrumental case study is used to provide insight into an issue; whereas an intrinsic case study can be used to gain a deeper understanding of the case; and a collective case study may be used in a study of a few cases to investigate a specific phenomenon. Personal interviews with important stakeholders, direct observation, psychometric tests, historical materials, and multiple site visits are all possible methods in case study designs (Lawrentz & Huffman, 2006)

Status and Survey Designs

Designs for Status and Surveys produce quantitative data, but they are frequently categorical and focus on opinions rather than hard outcomes. They facilitate the effective collection of perceptual and opinion data regarding evaluation objects, which are frequently the desired outcomes of the evaluation. Status studies provide point-in-time data on the prevalence of variables of interest in a given population (for example, frequency of computer use in schools or mathematics achievement). Status studies are significant because they establish benchmarks against which local evaluation findings may be compared (Lawrentz & Huffman, 2006).

Interpretive Designs

There are various types of interpretive designs used in STEM education. These types of designs are often used as formative evaluations where the opinions of people who are participating in the project are gathered and used to make inferences. They can use a variety of methods and methodologies: phenomenology, site visits, focus groups, and individual interviews (Lawrentz & Huffman, 2006).

A Site Visit Evaluation

Visits are made to select sites of multisite projects existing in local, state, or national settings. Diverse teams with expertise in evaluation, educational programs, educational administration, and business and industry conducted the visits, guided by prepared protocols. In an innovative approach to dissemination, the site visit reports could form the basis for a set of issue papers addressing important elements of a STEM program. These issues could range from collaboration, dissemination, materials development, program improvement, professional development, recruitment, retention, and sustainability (Lawrentz & Huffman, 2006)

Mixed Methods Designs

Most STEM education program evaluations use mixed methods to some degree. However, there are a variety of ways to mix methods. The major ones are the mixing of philosophies, research designs, and data collection devices. The mixing of philosophies implies the dialectic approach that Caracelli and Greene (1997) suggested. It is possible for the Mixing of research designs to be dialectical, but it is not required. One part of an evaluation could use one design while another part uses a different design, or the designs could be implemented sequentially. Furthermore, the results arising from these designs could be combined in various ways (Lawrenz and Huffman, 2004). The third approach, mixing data collection devices, is commonly used in STEM educational evaluation, and is frequently believed to increase the validity of findings or expand the scope; it can also include mixing of philosophies (Lawrentz & Huffman, 2006).

Mixing Philosophies

A Local Systemic Change (LSC) program evaluation serves as an example of a mixing philosophy (Horizon Research, 2005). The LSC program was designed to provide intensive

mathematics and science teacher education. In this program, each funded project was required to gather specific information using pre-designed evaluation instruments and could add its evaluation components. The data from all the projects were synthesized into program evaluation reports. The overall evaluation design was a purposeful mix of approaches. Projects observed educational programs and classrooms, furnishing numerical ratings as well as an interpretive explanation. Teachers, principals, and staff completed surveys about their opinions and behavior and were also interviewed (Lawrentz & Huffman, 2006).

Mixing Designs

A design experiment attempts to support arguments constructed around the results of active innovation and intervention in classrooms (Kelly, 2003). It is aimed at understanding learning and teaching processes in which the researcher or evaluator is active as an educator. The design varies as the innovation progresses, sometimes using positivistic stances and sometimes interpretive, as well as a variety of methods in pursuit of an effective and theoretically sound innovation (Lawrentz & Huffman, 2006).

Mixing Data Collection Activities

Is based on the philosophy of using a set of processes for collecting, analyzing, and "mixing" both quantitative and qualitative research on evaluation and methods in a single study to better understand a given research problem. One of the core tenets of STEM education evaluation is the improvement of student outcomes. There is also substantial evidence that, while the classroom learning environment is not the main predictor of student accomplishment, it is an essential moderator that educators can control. As a result, in STEM evaluations, improving the classroom learning environment is frequently cited as an objective (Lawrentz & Huffman, 2006).

In terms of the Discourse surrounding evaluation methodologies, the rationale that determines the choice of evaluators' methods was based on the type of questions that need to be answered. Regarding the choices that govern data collection, the case study design was the most cited. At least one dozen studies utilized the case study design to inform their research data collection methods. Six of the studies used critical reflection analysis and survey methodologies, respectively. The least used but credible method choices were observations, document analysis, online citations, and alternative presentation (including visual display, performance, multiple program theories, and poetry). Twelve papers comprising four theoretical and eight others were empirical and talked about focus groups as a means of collecting data. Four studies utilized interview techniques exclusively, and at least two of those studies used interviews in conjunction with other method choices.

Comparison of Approaches and Methodologies

The Values-Engaged, Educative (VEE) Approach

One of the top-cited approaches, the Values-engaged, educative approach has its roots in the framework for evaluation with a general Discourse on STEM education programs (Greene, et al., 2011). The VEE approach, while prioritizing value stances, has two main Discourses, one of which is to signal purposeful attention to the values intrinsic in education programs. The other Discourse of values engagement calls for evaluators to meticulously attend to the values of diversity and equity in each context. All the papers that cited VEE grounded their study or theoretical narratives on the "valuing" principles espoused by the VEE. Additionally, some of the articles (Boyce, 2017; Hall, et al., 2012) used a case study design as methods choices for data collection while another study (Johnson, Hall, et al., 2013), explored alternative approaches of

"Visual display, performance, multiple program theories, and poetry" to surface the values of its program participants.

Major Strength/Gaps in the VEE

The VEE evaluation approach (Greene, et al., 2011; Hall, et al., 2012) is better situated to address the needs of STEM education program evaluation, particularly in terms of NSF's Broadening Participation agenda (Boyce, 2017). It can strengthen credibility when attentively addressing issues of culture, diversity, and equity. Discourses in the VEE also encourage value stances such as trust and relationship building and could increase "knowledge regarding when, how, and to what extent" to attend to issues valued (Boyce, 2017).

One major gap in this approach is that it is time and human/resource-intensive: time, effort, skills, and expertise are needed to conduct the VEE approach. Additionally, because the VEE Discourse rejects objectivist evaluation, and instead subscribes to a postmodernist view, wherein there are no best answers and preferable values are prioritized, the program evaluation may culminate in conflicting findings and conclusions, leaving interpretation open (Stufflebeam, & Coryn, 2014) (see appendix A. for a summary of strengths and gaps in STEM education evaluation approaches and methodologies).

Systems Thinking/Oriented Evaluation Approach

Another top-cited approach, the Systems thinking/oriented evaluation approach is rooted in the traditions of theory-driven evaluation, logic modeling, and systems science and uses evaluation and program planning as the bridging mechanism between research and practice (Trochim, 2009). The main Discourse within this approach can be traced to addressing program evaluations that consider the complex factors that are inherent in the larger system within which programs are embedded. It can also be used in collaboration with other approaches such as the

case when Thomas and Parsons (2016) combined a culturally responsive evaluation and systems-oriented evaluation approaches in an evaluation capacity-building project. In that project, a group of STEM education evaluators focused on learning ways to shift their project evaluation toward a more systems orientation. The culturally responsive component of the project focused on negotiating solid relationships and boundaries and brought coherence to many systems concepts for practical application.

Major Strengths/Gaps in the Systems Thinking/Oriented Evaluation Approach

The conversations and debates within these approaches can illuminate important philosophical and theoretical distinctions in the evaluation and program planning that utilizes research and practice for theory-driven evaluation, logic modeling, and systems science (Urban & Trochim, 2009). The systems thinking approaches are useful when conducting program evaluation that considers the complex factors that are inherent in the larger system within which programs are embedded. However, it could be challenging to build a grounded theory that needs a thorough systematic, empirical process of observing events or analyzing materials drawn from operating programs, followed by an extensive modeling process.

The Citation Analysis (CA) Approach

In this approach, the Citation Analysis was frequently used within the Discourse of science research communities to measure the relative influence of scientific research enterprises and/or individual scientists but has limited use and impact in evaluation to date (Greenseid, & Lawrenz, 2011). In the citation analysis (CA), the evaluation can link up large-scale, multi-site STEM evaluations. It works well when conducting impact evaluation.

Major Strength/Gaps in the CA

The CA was useful for curating a large amount of evaluation data from a variety of web-based sources and search engines and can be used to compare the influence of multisite STEM education evaluations on the fields of education and evaluation by analyzing citation data (Greenseid, & Lawrenz, 2011). Some of its most effective methodologies use citations to access a list of evaluation reports, instruments, publications, and presentations produced by multisite STEM evaluations. However, the major gap or flaw of this approach is its heavy reliance on statistical, content, and network analysis. It has no capacity for direct human interactions, thus making the credibility of findings a major issue.

The Case Study Approach

The synthesis of information from multiple sources of evidence for triangulation, richness of data but with more variables of interest than data points, and analytic, rather than statistical, generalization of findings are all described by the case study technique (Simons, 2009; Stufflebeam, 2001; Yin, 1997). Due to the in-depth, relational approach between stakeholder involvement and evaluation use, concrete experiences can be achieved by sustained proximity to the observed reality and feedback can be obtained from participants under inquiry (Brandon, 2011; Flyvbjerg, 2006; Stufflebeam, 2001).

In essence, the case study method investigates a program's internal operations and how it integrates inputs and processes to produce outcomes in its geographical, cultural, organizational, and historical contexts (Stufflebeam, 2001; Yin, 1997). The case study approach distinguishes among central dominant themes, variations, and aberrations in single and multiple case studies. It identifies and describes the targeted and actual beneficiaries of a program evaluation. A case

study evaluates the needs of a program beneficiary and how well the program addressed those needs (Stufflebeam, 2001).

A Major Gap in Case Study Approach

Using the Case study design helps to focus the evaluation on an evaluand like a program or a small number of sites and can help to get to depth and richness of granular details; however, caution is advised because the findings are not always easy to generalize from a small number of cases. Additionally, case studies often rely heavily on qualitative methods to gather rich data that supports a thorough discussion of the site, which also limits the generalizability of findings.

Other Approaches and Methodologies

The following Discourses of evaluation approaches are not dominant in the field, but they appear distinct and tailored to specific populations within the context of STEM education programs. Among them are the Social Agenda/Advocacy, Responsive, Constructivist, and Deliberative-Democratic (Mertens & Hopson, 2006). This approach appears to pull many strands of philosophical underpinnings and is focused on making a difference in STEM education through program evaluation. The Discourses tend to focus on addressing issues of power, diversity, social justice, and human rights, including surfacing the voices of underrepresented groups in evaluation, and illuminating inequities based on gender, race, ethnicity, disability, and other relevant dimensions of diversity. One identifiable gap with these approaches is when an evaluator, intent on serving the underprivileged and empowering the disenfranchised in the community, cross over to comprise data to the benefit of the groups he/she is serving. Other approaches gleaned from the literature made use of a combination of citation analyses, online surveys, and surveys followed up with Interviews (Roseland, Greenseid, Volkov, & Lawrenz, 2011). In addition, there exists the Discourse of participatory/collaborative approaches in STEM

evaluation (Lawrenz and Huffman, 2004); and evaluation capacity building (Huffman, Lawrenz, Thomas & Clarkson, 2006). Methodologies used vary from Surveys (web-based and paper and pencil), Observation protocols, and Survey Design Used survey findings to follow-up, and engage in discussion and debate with colleagues about teaching science.

In summary, the approaches and methodologies used in most cases were determined by context, foci, and program community needs and expectations. Data collections from case study design to online/paper and pencil surveys, and visual/performance were driven by theoretical understanding and practical explorations that involve multiple levels of stakeholders' participation and engagement.

STEM Evaluators in Contexts and Settings

STEM education program evaluations address components that include how projects function within the economic, social, and political environment of its community and project setting, otherwise known as the context of the evaluation. More pointedly, the context of a program includes the environmental characteristics, assumptions, and external factors that may affect the implementation, replication, and generalizability of a given STEM education program. Evaluators in the field of STEM at single or multisite evaluation initiatives, K–12 schools and institutions of higher education, science outreach programs, and others that build capacity projects with goals to increase research in the disciplines of STEM education.

STEM education evaluators work in areas that are broadly inclusive of underrepresented groups; serving scientists, engineers, educators, students, and the public across the nation; and exploring every opportunity for partnerships, both nationally and internationally. They attend professional development workshops and meetings to present and share their work in places such as AEA STEM TIG where they focus on the application of evaluation to STEM programming as

well the exploration of the art and science of evaluation within the context of STEM. Through their efforts, evaluators hope to both improve the practice and status of STEM evaluation and enhance clients' use of evaluation approaches and findings.

In summary, this dissertation will use findings from the STEM evaluation literature as a basis for analysis of the abstracts and presentations extracted from the STEM TIG of the AEA conference. The study will follow up by interviewing STEM education evaluators and practitioners working in the field to understand the extent to which initial findings from document analysis resonate with their experiences. This study hopes to illuminate the perspectives of evaluators regarding the present and future states of STEM education evaluation.

CHAPTER III: METHODOLOGY

Introduction

In this chapter, I outlined the methodology used for this research. To begin, I restated the research objectives and followed up by describing the following components of the research methodology carried out in the study: (a) Research Design; (b) Research Ontological and Epistemological Framework (my standpoint as an educated, Black African man in America, with a STEM education background, who embraces a plurality of paradigms with a Social Justice lens); (c) Data Collection Methods; (d) Data Analysis; (e) Data Quality Criteria; (f) Data Collection Timeline; and (g) Research Design Limitations.

Research Objectives

The main aim of this study was to review and critique the Discourse of STEM Education Evaluation from the perspective of practitioners in the STEM education ecosystem. The central focus of the research seeks to understand the variety of Discourses surrounding evaluation approaches and methodologies of STEM education programs. More broadly, this study investigated the Discourse of current evaluation approaches undergirding philosophical and theoretical commitments of STEM evaluators and the widespread methodologies that determined choices of methods utilized in practice by STEM evaluators in different STEM education settings and contexts.

In this dissertation, I analyzed a sample of American Evaluation Association (AEA) STEM Topic Interest Group (TIG) evaluation abstracts to understand the Discourse of STEM Education Evaluation presentations that practicing STEM evaluators used, the sources of those presentations, the processes they used to choose their approaches and methodologies, and how they articulated their discursive contexts and settings. I based my presentation's analysis on the

types and sources of Discourse outlined by Mills (2004) and the ways evaluation approaches can be inferred in a STEM context (Greene, et al., 2011).

This research also contributes to a deeper discussion around evaluation approaches and methodologies prevalent in STEM education, and surface evaluators' current and future viewpoints of the STEM field with consideration for the Covid-19 pandemic. Additionally, this study contributes to the dialogue and other communication about STEM education evaluators' current and future perspectives of the field given the evolving Covid-19 pandemic.

To achieve the research objectives, the main research questions that guided this study are restated below:

1) What was the Discourse of STEM Education evaluation presentations at the annual American Evaluation Association conference from the year 2014 to 2019?
    a) What are the main evaluation approaches and methodologies utilized in these presentations?
    b) What are the differences between evaluators who work in Traditional Institutions versus practitioners in private organizations?
2) What are STEM evaluators' thoughts and impressions about the current and future state of the field?
    a) To what extent do the initial findings from the STEM evaluation literature of Discourses resonate with STEM education evaluators about the current state of the field?

Research Methodology

The research examines the Discourse of STEM Education Evaluation regarding approaches and methodologies utilized by practitioners in the field. This investigation was

conducted using a suitable qualitative research method where the researcher relied on the perceptions of STEM education evaluators' experiences (Stake, 2010) utilized in evaluation practice. This contrasted with quantitative approaches, where the researcher aims to better understand relationships among factors or variables within a given evaluation context using (typically) numerical data (Creswell, 2003).

Specifically, the research followed two distinct phases sequentially. Phase one was a qualitative study in which AEA conference abstracts were coded to examine how STEM education evaluators are engaging in Discourses relating to evaluation approaches and methodologies. I reviewed AEA STEM TIG presentation abstracts from 2014 to 2019 and conducted a content analysis. The second phase aimed to answer the second research question. I conducted 13 individual semi-structured interviews with STEM evaluators until saturation was reached, i.e., no new themes or information emerging (Richards & Morse, 2012; Strauss & Corbin, 1998). The individual interviews were key to better understanding practitioners' impressions or experiences about their work. Additionally, the interviews provided an opportunity to learn more about STEM education evaluators' thoughts on the current and future state of the STEM evaluation field. Among the emerging Discourses were evaluators' perspectives on funding, Broadening Participation, equity, and diversity issues in STEM education evaluation.

This sequential qualitative research, with initial content analysis of presentations at AEA STEM TIG followed by interviews of evaluators who are working in the STEM field, provided the ability to probe and triangulate findings and ask questions about the STEM education evaluation Discourse. The focus of investigations was (a) any similarities,
(b) any differences and (c) overlaps existing within the evaluation approaches and methodologies

deployed by STEM evaluators in practice. Having knowledge or understanding of these issues makes it valuable in the research design, and data gathered from content analysis benefitted the design of interview protocols that facilitated conversations during interviews with STEM education evaluators.

Research Design

The Discourse of STEM Education Evaluation regarding approaches and methodologies is steadily increasing, and this present research was seeking to understand the current or prevailing status of evaluation approaches and methodologies that are happening in conversations at conferences and within professional presentation spaces. To this end, a document review of presentations at the AEA professional meeting was conducted and followed up with interviews of STEM education evaluators in the field to gain insight into their perspectives on the Discourses surrounding the evaluation of STEM education programs. In this dissertation, I examined the Discourses surrounding approaches and methodologies used in the evaluation of STEM education programs. By examining the prevailing approaches and methodologies in these settings, my research seeks to improve knowledge of the STEM education evaluation Discourse. This study sought to understand what type of evaluation approaches and methodologies are discussed theoretically and utilized empirically in the literature and within STEM education professional spaces where practitioners meet to dialogue and advance the status of the evaluation field.

In this dissertation, a comprehensive literature review was conducted to explore different conceptions of approaches and methodologies used in the Discourse of STEM Education Evaluation, examined the current status within evaluation theory and practice, and highlighted areas of future research interests to practitioners. This served as the foundational material and

information utilized for the document review of AEA STEM TIG abstracts, and subsequent interviews conducted with STEM education evaluators.

The first question in this research was answered by conducting a review of the abstracts from AEA STEM TIG presentations available between 2014–2019. It was then followed by answering the second question, an empirical study, where I conducted 13 semi-structured individual interviews with evaluators who work in academic institutions and practitioners who work in private or governmental organizations within the STEM education field.

Ontological and Epistemological Framework

I used a variety of sequential data-gathering techniques including (a) a document review of the AEA conference material and (b) evaluator interviews to respond to my research questions. These two data-gathering methods are qualitative research methods with subjective epistemological and ontological assumptions. In this research, I espoused the democratic principles of a plurality of worldviews; embracing equity, diversity, inclusiveness; and consent of the research participants. I leaned on a social constructivist theoretical approach to help me comprehend my research topic considering my paradigm stance (Guba & Lincoln, 1994; Stake, 1995).

As a social constructionist researcher, I believe that multiple realities exist in ways that could be overlapping or intersect with one another. I believe these realities are created and constructed by the feelings, ideas, and opinions of the individual evaluators in the theory and practice of their profession. I deployed the constructivist lens in this dissertation to pay attention to the deliberation of the prevailing Discourses of STEM education evaluation. Previous studies (for example, Abdulwahed, Jaworski, & Crawford, 2012; El-Deghaidy & Mansour, 2015; Mertens & Hopson, 2006; Niess & Gillow-Wiles, 2013; Petrosino, Sherard, & Tharayil, 2020;

Radloff, & Guzey, 2016; Tyler-Wood, Cockerham, & Johnson, 2018) have drawn upon a social constructionist epistemology to surface various Discourses surrounding evaluation approaches and methodologies of STEM education programs.

The ontological stance that informed this research framework is relativism. As a research paradigm, relativist ontology does not believe in the existence of one *true* reality, and it does not prioritize a single *objective* truth. However, relativism subscribes to the reality that is *relative* according to how evaluators experience it at any given time and place.

Relativism believes that human sense-making is an act of construction from a conceptual system and independent of any foundational reality (Guba & Lincoln, 2001). Determining how evaluators' personal experiences influence their judgments of the present and future states of evaluation of STEM education programs is therefore at the heart of the constructionist research philosophy that served as the foundation for this investigation.

Due to the relativism-based philosophical stance that this investigation took, great care was taken to ensure that the information gleaned from conference abstracts was accurate, and interview analysis was done with the recognition that evaluators' reflections during the interview are relative to their experience domains or the contexts in which they work. Essentially, the adopted framework investigated the evaluator's interpretive work as subjective, weaving values, meanings, and facts, and prioritizing questions of quality and rationale of understanding, and was less concerned with questions of methods (Mertens & Hopson, 2006).

In summary, from an ontological perspective, I believe multiple realities exist in the Discourses of STEM education evaluation. The socially constructed diversity of dialogue drawn from evaluators' experiences facilitated the rich data gathered in this research. In addition, I adopted an epistemological perspective with the belief that STEM evaluators who participated in

my study have the knowledge and experience about the STEM education evaluation Discourse. This knowledge was uncovered through dialogue with STEM evaluators during individual, semi-structured interviews.

Phase One of the Research: Analysis of AEA STEM TIG Abstracts

Phase One of this research offered a full panorama of the STEM TIG abstract presented over six years (2014–2019) at the annual AEA conference. I presented a critique of the evaluation approaches and methodologies used and highlighted the underlying philosophies that undergird practitioners' Discourses on topical issues in STEM education evaluation using the information gleaned from evaluators' abstracts at the AEA. The follow-up study in Phase Two, which consisted of semi-structured interviews with STEM education evaluators, was informed by the findings from Phase One. Essentially, Phase One provided the conceptual framework that guided this investigation. In Phase One, evaluation approaches and methodologies that were most frequently used in conference presentations were identified, along with characteristics that set academic, corporate, governmental, and nonprofit organizational job profiles of evaluators apart.

Findings from Phase One contributed to the current conversation about STEM education by elaborating on a wide range of evaluation approaches and methodologies that are common in the domain of STEM. Additionally, STEM education evaluation is a relatively novel area within the evaluation community; thus, this study offered a good opportunity to advance research and evaluation knowledge in that field.

Documents Reviewed: AEA STEM TIG Abstracts

I conducted a systematic review of the AEA STEM TIG electronic abstracts. In addition to the STEM TIG conference documents, I examined and interpreted the content to generate

meaning and acquire understanding and further developed the knowledge base surrounding the Discourse of evaluation approaches and methodologies of STEM education programs. The content of the documents contained values espoused by STEM evaluators, textual belief systems, practices, and knowledge that constructs reality and provides shared ways of understanding STEM evaluators' perspectives and their evaluation work regarding approaches and methodologies.

The documents served as primary sources of data about Discourses surrounding STEM education evaluation approaches and methodologies and experiences of evaluators with the current and future state of the field. The evaluation documents were used as the main sources of information for discussions regarding the methodologies and evaluation approaches used in STEM education, as well as evaluators' experiences with the field's present and foreseeable future. I analyzed the documents using inductive coding to describe and organize stated approaches and methodologies from one discursive context to another (Fereday & Muir-Cochrane, 2006; Thomas, 2006).

The inductive coding technique was used for this study due to its bottom-up approach, which allowed codes to be derived from the data without regard to what they looked like and instead allowed narratives to emerge from the conference abstracts (Fereday & Muir-Cochrane, 2006). The exploratory nature of this approach enabled distinct ideas to emerge from the document analysis. I developed a document analysis log frame using spreadsheets with Microsoft Excel templates (see Table F21, in Appendix F) to focus my analysis, and each abstract/presentation was taken through an iterative process of skimming, reading, and interpreting (Bowen, 2009). In addition, I continuously reviewed my coding scheme and identified emergent patterns and themes in the data (Miles, Huberman, & Saldaña, 2014).

Summary of Documents Reviewed

Following the combination of inductive coding and constant comparative review of the abstracts described above, I used axial coding in thinking through and breaking down core themes that emerged from the abstract data. This thought process was used to generate 12 main themes including (a) Broadening Participation; (b) Challenges; (c) Equity, Diversity, Inclusion (EDI); (d) Effectiveness; (e) Funding; (f) Impact; (g) Implementation; (h) Improvement; (i) Lessons Learned; (j) Mentoring; (k) Professional Development; and (l) Successes. In addition, thematic analysis was used to further reduce the 12 items into four high-level themes to aid the presentation of results in chapter 4. Emergent themes in phase 1 of this document review were used to motivate and conduct interviews of individual STEM education evaluators in Phase Two of the research.

Phase Two of the Research: Interviews with STEM Education Evaluators

Phase Two of this investigation began by making use of the emergent themes from Phase One, the content analysis of the AEA STEM TIG abstracts. Findings from Phase One informed and guided the development of a semi-structured interview protocol used with STEM education evaluators in Phase Two. The research used the qualitative interviews carried out in this study to get in-depth information from participants on their experiences with STEM evaluation approaches and methodologies. Participants were recruited in two ways: through access to publicly available AEA Topical Interest Group LISTSERVs where over 40 members of the AEA STEM TIG were sent emails and by emails distributed to and shared by colleagues and some of the research participants.

In addition to the initial recruitment email, participants received one reminder email. All evaluators who expressed interest were sent a follow-up email to ascertain their availability for

interviews. Some were unavailable during the data collection period, resulting in their elimination from the study. The final number of participants was 12 evaluators from AEA; one participant was recruited outside of AEA membership, so the final data collection points consisted of 13 interviews.

In building on the findings of the content analysis of Phase One, the Phase Two study focused on how STEM evaluators characterized Discourses connected to evaluation approaches and methodologies utilized in their domain of practice. Some of the dominant evaluation approaches featured in dialogue during interviews with STEM evaluators include collaborative, utilization-focused, values-engaged educative, culturally responsive, and responsive ones. STEM evaluators used one or a combination of quantitative, qualitative, mixed-methods, scientific, and other methodologies.

Additionally, interviews conducted allowed participants' thoughts and impressions on the current state of the STEM field to emerge. The interview offered unique information and discovery of future ideas about STEM education evaluation Discourse regarding approaches and methodologies. To guarantee discussion of topics related to the subject matter while also affording freedom for further investigation of themes that emerged in each interview, a semi-structured interview style was used (Brinkmann, 2018; O'Leary, 2017; Patton, 2015). All the research activities for this study were conducted with the approval of the Institutional Review Board #: IRB-FY22-414, at UNCG at the end of November 2021. Table 3 below shows each data collection method used for this study

Table 3. Data Collection Methods, Research Questions, and Methods of Collection

Data Collection Method	Phase One			Phase Two	
	Q1.	Q1a.	Q1b.	Q2.	Q2a.
	What was the Discourse of STEM Education evaluation presentations at the AEA from 2014–2019?	What are the main evaluation approaches and methodologies utilized in these presentations?	What sectors and contexts are STEM evaluators working in?	What are the STEM education evaluators' thoughts and impressions about the current and future state of the field?	To what extent do initial findings resonate with STEM evaluators about the current state of the field?
A). Content analysis of AEA STEM TIG Abstracts.	✓	✓			
B). Semi-structured individual interviews with STEM Evaluators		✓	✓	✓	✓

The Interview Protocol

For phase two of the research, the second data collection method was interviews with STEM education evaluators including those in academia, the private, public (or government), and other non-governmental organizations. The purposes of these interviews were to:

(a) Determine the approaches and methodologies that support evaluators' use of various philosophies and the justification for their theory and practice in STEM disciplines and across program domains of evaluation practice.

(b) Identify and examine evaluators' rationale behind the choice of methods concerning the prevailing Discourse of STEM Education Evaluation.

(c) Explore alignment between emerging Discourses from conference presentations and perspectives garnered from evaluators' conversations during interviews.

(d) Examine evaluators' approaches and methodologies for reflecting on current and future states of STEM education evaluation.

The full interview protocol (see Appendix B) was developed from the document review and content analysis of the AEA STEM TIG abstracts and in conjunction with the overall feedback from my dissertation committee, who ensured the quality of the textual information represented in the instrument.

Participant Selection and Rationale

I used a non-probability (non-random) sampling technique in this research. Specifically, I used purposive sampling, where participants were selected based on availability and willingness to take part in the study (Creswell, 2014). Thereafter, when it became apparent that potential participants were hard-to-reach groups, those who had participated were requested to suggest additional evaluators they knew to take part in the study. The purposive samples for this study

were recruited through available members of AEA STEM TIG whose abstracts were reviewed in Phase One of this research. Subsequently, the researcher sought the help of AEA STEM TIG evaluators to nominate potential participants known to them who work across multiple sectors and with varied priorities, such as PK–12 education, higher education, and governmental and private organizations (Malyn-Smith, 2014) to participate in the study.

It is noteworthy that, over the six years considered for this study, only four of the research participants, 2 from the academic community and 2 from the private sector, had their AEA STEM TIG abstracts identified and reviewed before meeting with them for interviews. There are several reasons for examining research participants' AEA STEM TIG abstracts before interviewing them for this study and these include.

Familiarization with the participants' work: By reviewing the participants' abstracts, the interviewer can gain a better understanding of their research interests and expertise. This familiarity can help the interviewer ask more informed questions during the interview and allow them to engage in more meaningful discussions with the participant.

Increases efficiency: Reviewing abstracts can help the researcher to develop a clearer understanding of the participant's work, which can help to facilitate the interview process. This can help to reduce the amount of time needed to conduct the interview and can help to ensure that the interview remains focused on the relevant topics.

Facilitates a more targeted interview: Reviewing the abstracts can help the interviewer identify potential areas of interest or concern that they may want to explore further during the interview. This can help to ensure that the interview covers all relevant topics and leads to a more focused and informative discussion.

Enhances quality of data: By reviewing the abstracts of potential interview participants, researchers can gain an understanding of the participant's perspectives and approaches to their research. This can help to ensure that the questions asked during the interview are relevant and likely to produce high-quality data.

Overall, reviewing the AEA STEM TIG abstracts of interview participants before meeting with them provided a valuable tool for me, the researcher, to gain a better understanding of their potential contribution and ensure that the interview is more focused and informative toward my research topic.

The study focused on the STEM education evaluators and explored their viewpoints regarding the current and future states of the field. The research shed light on how STEM education evaluators are experiencing evaluation practice at present and in what ways their experiences are shaping the future of the STEM education evaluation field. Further, the inquiry probed and uncovered the extent evaluators' perspectives resonate with preliminary evaluation literature findings.

Response Rate

The Senior Coordinator of programs for AEA was contacted to provide direction to the publicly available conference websites or links where STEM TIG abstracts were located. Abstracts and presentations from the years 2014–2019 were identified and downloaded to a folder on UNCG BOX, a secured, cloud-based content management, collaboration, and file-sharing tool. Using information gleaned from the abstracts, AEA members who belonged to the STEM TIG were randomly selected and contacted via email to seek their participation in the study. Recruitment emails inviting STEM evaluators (see Appendix B) to participate in the study

contained a link to a consent form in Qualtrics (a cloud-based statistical management software platform).

A total of 40 emails were sent to STEM evaluators (those whose abstracts appeared in the years 2014–2019); 11 responded and consented; nine responded but declined; 19 did not respond at all, and one bounced back. The eleven who consented were redirected to an online calendar meeting platform to schedule availability to be interviewed on ZOOM. Two additional STEM evaluation practitioners were contacted based on recommendations from some of the consenting AEA STEM TIG members. The additional two STEM evaluators agreed to participate in the study, thus increasing the total number of participants to 13. Out of the 13, n = 3 were men and n = 10 were women. Further, n = 6 identified themselves as working in academia, and n = 2 identified themselves as working in both academia and the private sector; n = 4 were working in private organizations, and n = 11 worked in government. All participants were highly educated: the academic evaluators all had doctorate degrees (n = 6); four of the five evaluators who work in the private sector have a doctorate, and one has a Master's degree. Collectively, interview respondents had over 150 years of experience conducting STEM education evaluations and had conducted more than 90 evaluation approaches and methodology assignments.

Individual Interviews with STEM Evaluators

I conducted semi-structured interviews for an average of 75 minutes with a diverse group of evaluators working in the STEM ecosystem, ranging from novices to seasoned practitioners. These interviews explored a conceptual understanding of the Discourses of STEM education evaluation and illuminated the experiences of evaluators within the context of evaluation approaches and methodologies. As necessary, the results of the document analysis served as a guide during the interview. I had a one-on-one engagement with individual participants, probing

their experiences and getting a deeper understanding of their perceptions about the current and future states of the STEM education evaluation field. The individual interviews were focused, and more time and attention were devoted to exploring each evaluator's perceptions and experiences in much greater detail. I had more time to discuss and explore topical issues in detail and worried less about the group dynamics that often occur in focus groups (Babbie, 2010).

The goal of the follow-up interviews was to facilitate a more in-depth conversation where participants had the opportunity to explore and explain their thoughts in collaboration with the researcher. The interview protocol was created following the completion of the document review and provided a chance to get participant perspectives on important findings, discuss divergent or convergent ideas regarding emergent Discourses, and go deeper into the study topics. The chair and two members of my dissertation committee helped me establish the interview protocol (see Appendix B).

Interview Process

Interviewing subject matter specialists and practitioners is a method for fostering discussion, according to empirical studies of evaluation (Christie, 2003; Fitzpatrick, Christie, & Mark, 2009). A racially and culturally diverse pool of 40 STEM evaluators with different ethnic backgrounds, nationalities, sexual orientations, gender identities, sectors, ages, and years of practice in STEM education evaluation were invited to participate in a 45- to 60-minute telephone or online ZOOM interview. At the end of the process, the researcher succeeded in interviewing 13 evaluators. The interviews provided an opportunity for gathering information on how evaluators engage with the Discourse of STEM Education Evaluation through approaches and methodologies and the extent to which they perceive the current and future states of the field. I contacted STEM evaluators who are AEA STEM TIG members via email. The emails

included information about the purpose of the study, topics of discussion, procedures for the study (i.e., audiotaping, transcription of data, and presentation of the data), maintenance of confidentiality, and security of data. Each interview began with obtaining informed consent using a downloaded template of UNCG IRB with modifications to fit the use of my study.

A guiding interview protocol of open-ended, descriptive, and interpretive questions was developed in the context of STEM education evaluation to encourage flexible discussion with evaluators and assure the ethical properties of the data acquired. The interviews were audio recorded and submitted to Otter.ai, a speech-to-text transcription and translation application software. In one instance, the researcher honored the wish of a participant who declined to be recorded or videotaped. Recordings were transcribed in full, and transcripts were stored on a secured, password-protected laptop belonging to the researcher for onward analysis. After the study had been completed and the findings documented, the digital recordings of the participants' voices were destroyed. Information collected during the study was strictly confidential and used solely for research purposes. The transcriptions were checked for accuracy against the audio and corrected before coding. Handwritten notes were taken during the interviews, and periodic member checks of the notes were conducted with the participant.

Each of the interviews began with a reminder of the main goal of the study, which is to examine STEM education evaluation Discourses regarding the approaches and methodologies utilized by evaluators within the STEM ecosystem. The focus of the study was on how STEM evaluators characterize Discourses connected to a gamut of evaluation approaches and numerous methodologies that justify the methods used in practice. Specific examples of evaluation approaches and methodologies were given as conversation starters; however, a careful and deliberate effort was made to recognize how to interview respondents may embrace a different

terminology or choose not to adopt a particular evaluation approach or methodology, and allowance was made to accommodate a broad range of practices or ideas about how STEM evaluators conceptualize their work and operate in practice.

The participants were then verbally asked to confirm their willingness to have their comments audio-recorded, allowing me to preserve their fidelity and integrity as well as for transcription purposes. Since the study focuses on how STEM evaluators characterize Discourses connected to the panorama of evaluation approaches, a deliberate effort was made by the transcriptionist to highlight the term "evaluation approaches' to cover a broad range of practices or ideas about how to conduct STEM education evaluation. Consequently, I made allowances to accommodate interviewee plurality and diversity of expression that align with the context of this study. Furthermore, I started each interview by referencing methodologies as fundamental research strategies that explain the logic of research methods. Through dialogues and in the analytical process, I made sure to emphasize evaluators' theories and praxis. When requested, the interview script was emailed to the participants so they could review it ahead of schedule and have any questions or concerns addressed by the researcher.

Data Synthesis and Analysis

Phase Two data were first prepared for analysis by transcribing the online ZOOM recorded interview conversations into a text format using otter.ai. This allowed for ease of data coding and the opportunity to become more intimately familiar with the data. In completing the data analysis of the emergent Discourses of STEM evaluators' interviews, I started by constructing an analytic matrix (Miles, Huberman, & Saldaña, 2014) on a Microsoft Excel spreadsheet that displayed the perspectives I had derived from the AEA STEM TIG abstracts and presentations on rows horizontally across the top of the matrix, then displayed each distinct

perspective of STEM evaluators interviewed vertically down the side of the matrix. I then placed a brief excerpt of data (containing evaluation approaches or methodologies) in the cells of the matrix to indicate which Discourses resonate with the thoughts and impressions expressed by each interviewee.

I also placed brief excerpts in the cells to indicate any Discourse that was identified as particularly insignificant from a specific perspective. I highlighted cells that indicate agreement in findings between perspectives and used a different color to highlight cells that indicate a conflict between perspectives. This matrix served as the basis for analyzing the extent of resonance across the varying perspectives expressed by STEM evaluators and for constructing the full picture of the current state of the STEM field in terms of approaches and methodologies, including the future of the STEM education evaluation enterprise. The constructivist-grounded theory utilized in the interview analysis enabled an iterative and comparative process where the researcher analyzed data from earlier interviews before conducting the later ones (Charmaz & Belgrave, 2012).

Consequently, the synthesis of data occurred iteratively as new sources of data were added per interview (Charmaz & Belgrave, 2012), which informed the comparative interview analysis, and as complete interview synthesis occurred (informing the cross-interview analysis). This process was, therefore, iterative, and ongoing throughout the study, and allowed for the most thorough and accurate exemplification of each interview and to identify nuances between individual interview data.

Thematic content analysis was used to examine the interview data holistically and categorically, detecting patterns across individual interviews and placing those patterns within the contexts of each interviewee's experience. As a result, the study used the constant

comparative method to identify themes. The constant comparative method is an inductive method where data are "analyzed and split into codes based on emerging themes and concepts, which are then sorted into categories that represent an analytic comprehension of the coded entities' (Glaser, 1965). A thorough memo-writing method was used to assist this process during the study phase (Strauss & Corbin, 1998).

During the coding process, constructivist grounded theory (Charmaz, 2014) influenced how I interpreted the data. I read and reread the data using the assumptions delineated by Charmaz's framework. Charmaz and Belgrave (2012) highlighted the assumptions that (a) multiple realities exist, (b) data reflect researchers' and research participants' mutual constructions, and (c) the researcher enters, however incompletely, the participant's world and is affected by it.

The major influencing assumptions of this qualitative inquiry that impacted the data analysis were that the Discourses of STEM Education Evaluation can be surfaced or reflected through seeking multiple perspectives and experiences of evaluators within the field. In addition, analytic attention was paid to the language used and the values implicit in the evaluator's experiential views of evaluation approaches and methodologies. Together, these two frameworks helped concretize the emergent Discourses of STEM Education Evaluation as constructions of realities expressed by evaluators for the present and foreseeable future of the field.

The major influencing assumptions of this qualitative inquiry that impacted the data analysis were that the Discourses of STEM education evaluation can be surfaced or reflected through seeking multiple perspectives and experiences of evaluators within the field. In addition, analytic attention was paid to the language used and the values implicit in the evaluator's experiential views of evaluation approaches and methodologies.

Together, these two frameworks helped concretize the emergent Discourses of STEM education evaluation approaches and methodologies. Specifically, I conducted a constant comparison of individual interviews and identified emergent themes from the interview data that reflected these assumptions, including seeking to understand how evaluators' expressions of realities influence their current and future perceptions of the STEM education evaluation field. I sought to hear the evaluators' full story during the interviews while probing for the analytic properties and implications of major narratives during the analysis and synthesis processes. Therefore, there was an interplay between analytic inductive processes (identifying themes that arose from the data) and constant comparison processes (identifying emergent themes through the assumptions of the theoretical frameworks). Furthermore, I created memos, reflected on the data and themes, and looked at the data to either confirm or contradict emergent stories or accounts.

Before coding and after transcription, all interview transcripts were reread. Then, the text was inductively coded line by line using MaxQDA software (2020). The first round of codes was reduced by identifying codes that were nearly the same and collapsing them. In the second round of data reduction, these codes were then further reduced in number. Bringing these codes into dialogue with my memos, the research questions, and the transcripts themselves, multiple overall themes were generated, linked directly to the qualitative research questions. Also considered during this process were original codes that appeared frequently among research participants, or that were particularly salient concerning Research Questions One and Two. This final group of findings was then used to revisit the original data to determine if they made sense in the context of the interviewees' original words.

The goal of this analysis was to surface the issues and identify patterns most relevant to the research questions. Chapter four will focus on those themes, supported by a selection of textual examples drawn directly from the transcripts. During analysis and interpretation, the constructivist grounded theory and constant comparative methods enabled findings to be connected to broader findings under consideration to gain common understandings around the Discourses of STEM education evaluation regarding evaluation approaches and methodologies.

Data Quality Criteria

In general, criteria that consider the social dimension of the investigation and the researcher's participation in the research context are used to assess the quality of qualitative research. Specifically, data quality in my study was established through "trustworthiness criteria," conceptualized as confidence in the accuracy and representation of findings achieved through credibility, dependability, confirmability, and transferability (Lincoln & Guba, 1985).

Credibility

Credibility generally describes how comparable to internal validity is the likelihood that credible findings and interpretations will be produced (Lincoln & Guba, 1985). The credibility criteria suggest that the claims made should be supported by an adequate amount of data (Stake, 2005) and that the analysis and interpretation processes should be made explicit (Mertens, 2015). I used triangulation, which primarily consists of a repetitive process of participant interviews followed by a critical review of what was being said by interviewees to reduce bias and gain a better understanding of conversations (Maxwell, 2013), thus increasing the overall trustworthiness of the findings (Stake, 2005). This approach is crucial when considering pluralism and the constructivist framework that underpins this study.

Multiple Methods

The utilization of several techniques to gather data (such as document analysis and interviews) improved the credibility of my findings (Taber, 2008).

One limitation of using document analysis of abstracts in Phase One is that the scope of research is invariably limited, as I was only able to examine publications for six years. This means that if information important to my study was left out, it would not have been available to provide input into my investigation, analysis, or, ultimately, any conclusions or findings that were reached. Furthermore, the opportunity to elaborate on and clarify evaluators' thoughts and impressions on the current and future state of the STEM education evaluation field, as well as to delve deeper into discussions of evaluation approaches and methodologies employed within the field offered by the interview with academics and practitioners, might have derived richer data had this information been available.

A probe-based interview protocol was developed, refined in conjunction with the content analysis of abstract data from Phase One, and deployed to obtain comments from interviewees in Phase Two. The interview protocol strengthened the focus and scope of the conversations (Stake, 2005).

Memoing

This was incorporated during the interviewing process to reflect on the process throughout this study (Treharne & Riggs, 2017). I kept a methodological journal and wrote reflections about my methodological journey. I used some of the journal's content to write short memos where my musings addressed the issues I was grappling with in this research. I had time to question the data and wonder to what extent I was seeing things and interpreting the information because of my academic background as a STEM major and/or my STEM evaluation

training. I raised these questions and weighed the interviews in writing the memos. My view of constructivist grounded theory encouraged me to treat the memos analytically and define emerging themes by their empirical properties. When I juxtaposed the interview data with the memos, I began to have a clearer picture of some of the issues I grappled with in the study. The process of contemplating, organizing, and eventually writing about my findings helped me develop emergent themes.

Transferability

Transferability is the degree to which the conclusions reached in research can be applied to other entities or settings (O'Cathain, 2010). By defining the theoretical underpinnings of the study and providing adequate context for quotations and conclusions, transferability helps the audience of this research to gain understanding. Explicit reporting is then used to promote transparency and enable readers to interpret transferability.

Dependability

According to Guba and Lincoln (1989), dependability, the postpositivist term for reliability, indicates that results hold up over time (Mertens, 2015). The implications of this criterion apply to having a credible inquiry process such that conclusions drawn from the study are reliable, rather than the usual connotation of whether the study can be reproduced. This is the foundation for transparency of my research strategy and how findings were informed. The diversity and inclusivity of the evaluator voices, who are experts in the STEM evaluation field, further secured the dependability of the study.

Confirmability

Confirmability is like the objectivity of the postpositivist paradigm (Guba & Lincoln, 1989). The concept of confirmability indicates that the data and their interpretation are supported by evidence and can be traced back to their sources, as can the reasoning behind how the interpretation was produced; however, the concept of objectivity denotes that the researcher's influence is minimal (Merriam, 1988). As stated earlier, the researcher's bias in the interpretation was limited by the triangulation of the data. The multiple data collection sources—254 abstracts, 94 pages of interview transcripts, and researcher journals made it simpler to build a confirmability check. This is because each piece of data that was woven into the interpretation could be traced back to its source, and the basis for the claims made was evident. Most of the data collection for this research took place between December 2021 and August 2022. Table 4 below displays the research data collection timeline.

Summary of Chapter Three

In sum, I carried out sequential qualitative research involving a multi-stage research method that combines two or more qualitative research techniques. Phase One involved an initial content analysis of the publicly available abstract documents from the AEA conference STEM TIG, while Phase Two involved individual interviews with evaluators who work in the STEM education evaluation field. In Phase One, I conducted a content analysis of the abstracts to identify themes and patterns related to STEM education evaluation. This provided me with a preliminary understanding of the topic and helped to identify potential participants for the individual interviews in Phase Two.

Table 4. Research Data Collection Timeline

Steps in the Research Process		Dec.2021–Jan. 2022	Jan.–Feb. 2022	Mar. 2022	Apr.–Jun. 2022	Jun.–Aug. 2022	Sept. 2022–Feb. 2023	March. 2023
1	Conduct document review	✓						
	Analyze data	✓	✓					
	Write up findings from Phase One		✓	✓				
2	Conduct Interviews with STEM Education Evaluators				✓			
	Analyze Data				✓	✓		
	Write up findings from Phase Two					✓	✓	
3	Complete Dissertation						✓	
4	Defend Dissertation							✓

In Phase Two, I conducted individual interviews with individual evaluators to gather in-depth information about their experiences, opinions, and perspectives on STEM education evaluation. This information was used to complement and expand upon the findings from Phase One, providing a comprehensive picture of the research topic. The value added to my research by using a sequential qualitative research design is that it combines the strengths of qualitative document analysis and individual interview methods to provide a more comprehensive understanding of the topic. It also allows for the refinement of my research questions based on the findings of the initial content analysis.

Research Design Limitations

A few limitations are advanced for consideration in this study. First, it is possible that, as the sole researcher, I may be biased in my appraisal of the Discourse surrounding the evaluation approaches and methodologies of STEM education programs. Rigorous research design, multiple data collection methods, high-quality criteria for data, and internal and external reviewers were employed to lower the effects of researcher bias.

Second, the timing and prevailing climate of the ongoing Covid-19 pandemic during which this study occurred are potential limitations. Rather than conducting the preferred, in-person individual interviews, participants felt comfortable virtually, but there were risks that internet or Wi-Fi connectivity became problematic. In such scenarios, interviewees were asked to reflect retrospectively to answer some of the interview questions but declined and rather waited for internet connections to be established before the interview process resumed and was subsequently completed. Lastly, as much as possible, to ensure the accuracy of interviewees' responses before interviews, I requested interviewees to review previous evaluation abstracts and presentations, evaluation reports, and evaluation plans

CHAPTER IV: FINDINGS

The Discourses of the STEM Education Evaluation Field

This chapter provides a detailed description of the findings of this research. Findings are presented for each of the two questions under investigation. The steps for presenting the findings follow. These include (a) presenting the research question, (b) defining and providing conceptual clarity where necessary, and (c) acknowledging the data that support the emerging category/theme. The next step was (d) presenting themes within the overarching finding, (e) expanding on each theme with the presentation of pertinent synthesized data, and then (f) offering an overarching finding for the questions asked. When switching from one theme to the next, the reader will be informed of the research question being addressed as well as the theme that is being developed. Finally, the findings will be communicated in a way that protects the confidentiality and privacy of conference abstracts, presentations, and research participants.

Findings by Research Question

Question 1: What was the Discourse of STEM Education Evaluation presentations at the annual American Evaluation Association Conference from the year 2014 to 2019?

- Overall, lessons learned, challenges, and impact were the most frequently mentioned constructs in the Discourse of STEM Education Evaluation by professionals in academic, private, and public sectors.

The research findings in this section provide a summary of AEA abstracts related to STEM education evaluation, comparing the debate, discussion, and dialogue on topical issues that took place in the six years (2014–2019) of the STEM TIG at the AEA annual conference.

Definition of Terms

- **The Discourse of STEM Education Evaluation** within the context of the AEA conference in this study is defined as ways of establishing or maintaining relationships among individuals in social and professional circles by employing language, ideas, moral principles, deeds, encounters, and exploring topical issues in specific spatial and temporal dimensions of conference format.
- **Academic Sector**. Include public or private institutions of higher learning that award academic degrees, and public or private non-profit research institutions whose main goal is to conduct research, and discharge duties with related goals.
- **Private Sector**. The portion of the economy that does not belong to the government is characterized by an ability to create for-profit or nonprofit organizations.
- **Public Sector.** This portion of the economy is composed of both public services and public enterprises that are controlled by the government.
- **Program Development** means the ongoing systematic process of planning, implementing, and evaluating a program, which involves a collaborative relationship between the Contractor and Administrator.

Overview of Abstract Data

Data used to answer this question included an analysis of AEA STEM TIG abstracts and presentation documents from the years 2014–2019. In the six years from 2014 through 2019, a total of 254 abstracts were identified for presentation at the STEM TIG meetings at the AEA conference (Table 5). A total of ten session types were identified, including (a) multi-paper, (b) roundtable, (c) panel, (d) think tank, (e) expert lecture, (f) skill-building workshop, (g) birds of a feather, (h) Ignite, (i) demonstration, and (j) poster. Overall, the year 2014 had the least number

of abstracts accepted, with eight (n = 8), and the year 2015 recorded the most acceptances, with 67 abstracts. Total abstract acceptance decreased to 39 in 2016, and production increased by 4% in 2017 and remained relatively stable at that rate for the next two years in 2018 (n = 47) and 2019 (n = 50).

Table 5. AEA STEM TIG Abstracts, 2014–2019: Session Type

Session Type	2014	2015	2016	2017	2018	2019
Multi-paper	1	31	13	13	17	24
Roundtable	2	2	5	6	6	4
Panel	5	13	5	7	7	9
Think Tank		3	1	1	1	1
Expert Lecture		1				2
Skill-Building Workshop			1	1	1	1
Birds of a Feather		1	1	2	2	1
Ignite		3	1	1	1	1
Demonstration		2	3	4	4	
Poster		11	9	8	8	7
Total	8	67	39	43	47	50

While multi-paper sessions remained number one, panel and poster sessions followed as the second and third most patronized sessions by STEM evaluators. Abstract session types with low acceptance include a demonstration with 13; think tank, birds of a feather, and ignite with seven each; skill-building workshops with four; and expert lecture with just three in total. Specifically, the year 2015 had the highest production with 31 multi-paper submissions. Collaborative followed in 2019 with 24, and in 2018 with 17 accepted abstracts. The years 2016

and 2017, which each had 13, and 2014, with just one, had the least productive years in terms of multi-paper sessions.

While the need for multi-paper sessions steadily increased over the six years under investigation, the number of roundtable sessions remained consistently low over the same period (75% less). Table 5 shows that multi-paper sessions consistently led in the breakdown of the AEA STEM TIG abstract session types over the six years (2014–2019) under review. Furthermore, the number of presenters at the AEA conference, STEM TIG, has been summarized and shown in Table 6. The abstracts for the presentation were grouped into either of these three: academic, private, or public. These three sectors were used to organize the findings for this research question.

The research used the systematic review as a framework to generate meaning and acquire an understanding of the Discourse of STEM Education Evaluation approaches and methodologies. The following themes emerged from the data analysis and bolstered the main findings in support of the central idea: *The Discourse of STEM Education Evaluation Presentations at the Annual AEA Conference from the Years 2014 to 2019.*

Insights across the three sectors revealed that the most frequently Discoursed of the STEM Education Evaluation presentations at the annual AEA conference from the year 2014 to 2019 in abstract conversations among the STEM evaluation professionals were (a) lessons learned; (b) challenges; and (c) impacts. Moderately mentioned items of discussion in the abstracts were (a) improvement, (b) implementation, and (c) effectiveness.

Table 6. AEA STEM TIG Abstracts, 2014-2019: STEM Evaluator by Sector

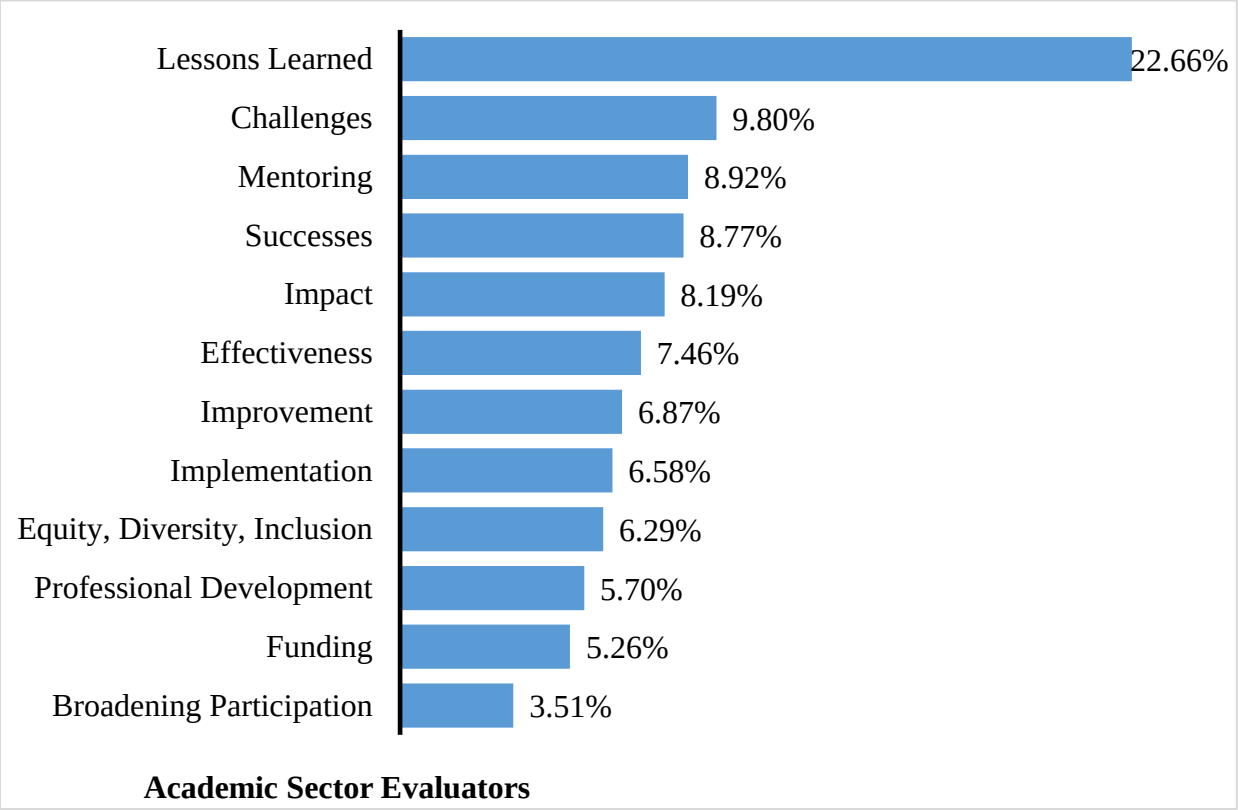
Year/Sector	Academics	Private	Public	Total
2014	5 (63%)	2 (25%)	1 (13%)	8
2015	41 (61%)	21 (31%)	5 (7%)	67
2016	25 (64%)	12 (31%)	2 (5%)	39
2017	29 (67%)	13 (30%)	1 (2%)	43
2018	26 (55%)	16 (34%)	5 (11%)	47
2019	25 (50%)	23 (46%)	2 (4%)	50
Total	151	87	16	254

The least frequently mentioned items that showed up in the abstracts include: (a) professional development; (b) equity, diversity, and inclusion (EDI); and (c) broadening participation (BP). Other issues, such as funding, mentoring, and successes, showed up in all the conversations among the STEM evaluators; however, to a lesser degree (see Table G22 in Appendix G).

Specifically, findings showed that academic sector evaluators most frequently mentioned lessons learned (22.66%), challenges (9.80%), mentoring (8.92%), and successes (8.77%). Items such as impact (8.19%), effectiveness (7.46%), improvement (6.87%), and implementation (6.58%) were moderately mentioned in the academic sector abstracts. The least-mentioned items in the academic evaluators' abstracts include equity, diversity, and inclusion (6.29%) issues,

professional development (5.70%), funding (5.26%), and broadening participation (3.51%). See Figure 1 below for the dominant Discourse.

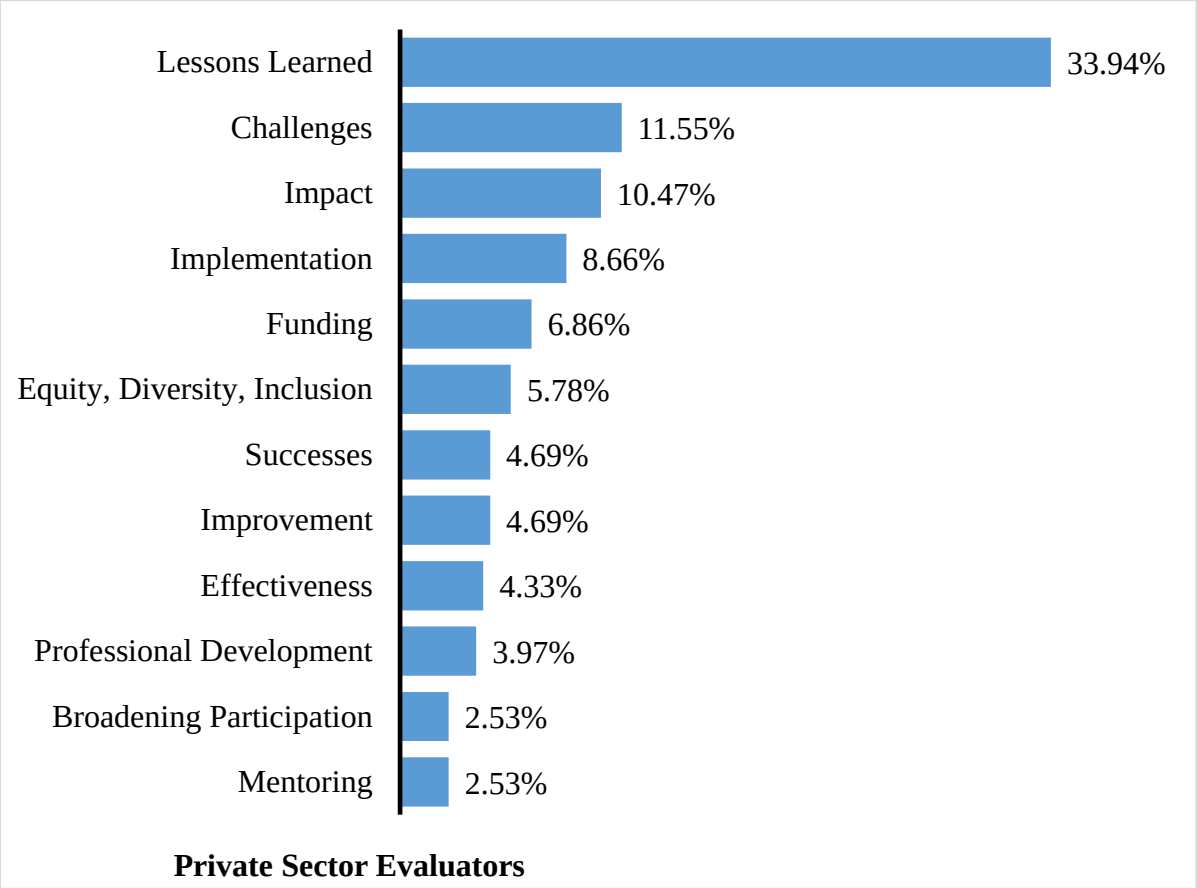
Figure 1. Major Items of Discussion in the AEA STEM TIG Abstracts by Academic Sector Evaluators



Following a similar pattern, private sector evaluators most frequently mentioned lessons learned (33.94%) and challenges (11.55%) in the discussion, except they also mentioned impact (10.47%) and implementation (8.66%) rather than mentoring and successes as the third and fourth most frequently discussed items in abstract conversation. Funding (6.86%); equity, diversity, and inclusion (5.78%); successes (4.69%); and improvement (4.69%) were moderately mentioned items in the private sector abstracts. The least frequently mentioned items by private

sector evaluators in the discussion were effectiveness (4.33%), professional development (3.97%); broadening participation (2.53%); and mentoring (2.53%). Figure 2 below show the Dominant discussion items.

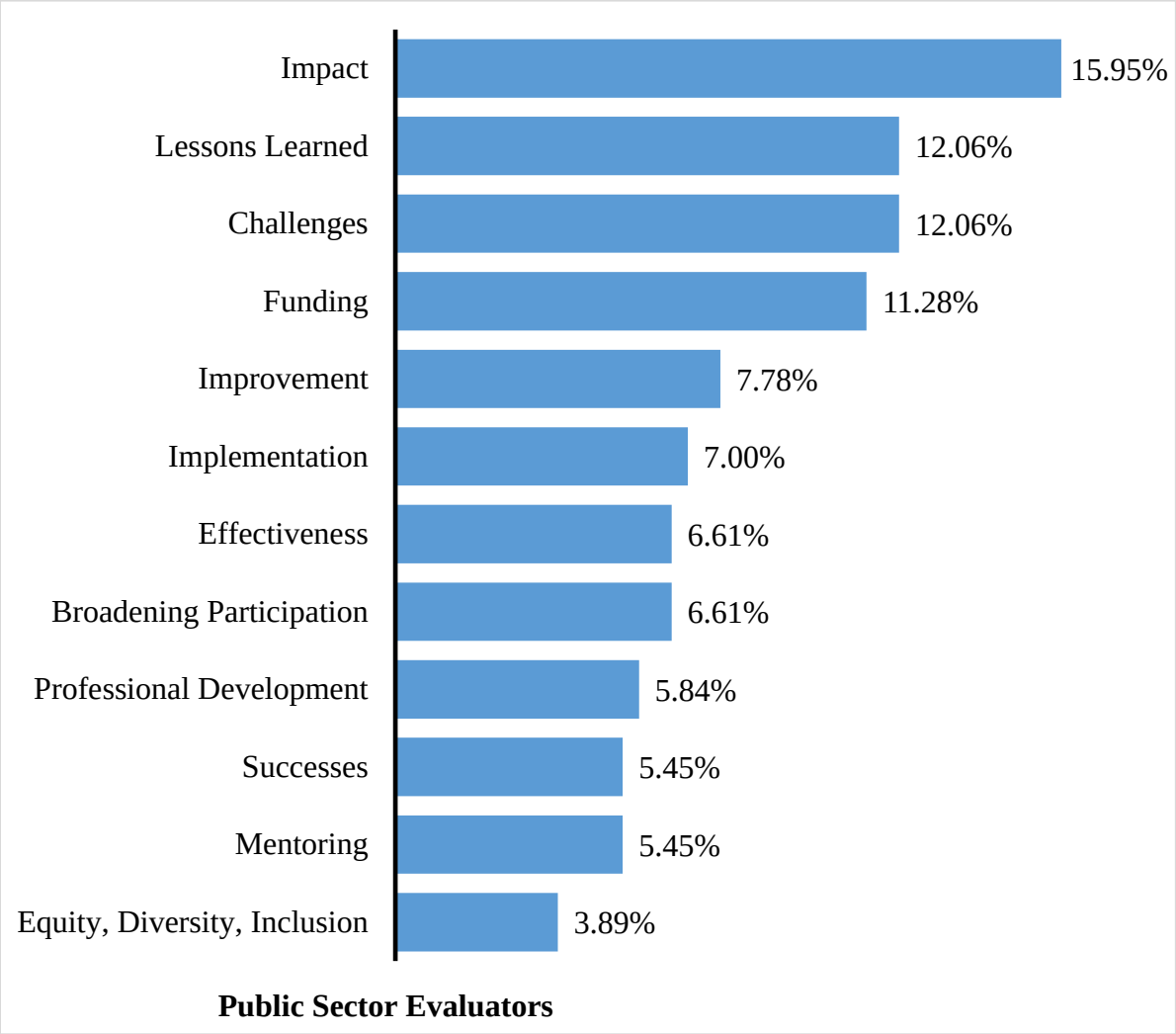
Figure 2. Major Items of Discussion in the AEA STEM TIG Abstracts by Private Sector Evaluators



In some deviations from the academic and private sector top four abstract findings, public sector evaluators most frequently mentioned impact (15.95%), followed by lessons learned (12.06%), challenges (12.06%), and funding (11.28%). The public sector evaluators moderately discussed improvement (7.78%), implementation (7.00%), effectiveness (6.61%), and broadening participation (6.61%). They paid the least attention to issues of professional

development (5.84%); success (5.45%); mentoring (5.45%); and equity, diversity, and inclusion (3.89%). See Figure 3 below for major items of discussion.

Figure 3. Major Items of Discussion in the AEA STEM TIG Abstracts by Public Sector Evaluators



Focused coding was conducted and used to organize all the main items into four major themes of Discourse, including (a) lessons learned, (b) broadening participation, (c) use of STEM education evaluation, and (d) program development. Sub-codes were identified for each thematic item; however, the variety of issues discussed made it necessary to structure the subcodes under the following four headings:

(a) Program/Project or Funding/Agency

(b) Evaluation Settings

(c) Program Group/Community

(d) Dominant Conversation

Additionally, each of the abstracts reviewed was classified according to the sector of practice in which the evaluation was considered to have taken place.

The Discourse of STEM Education Evaluation at the Annual AEA Conference

The following four themes emerged from the data after careful content review and analysis of the data.

Theme 1: The Presentations heavily focused on Learning Experiences in STEM education evaluation contexts.

Theme 2: Professional Development was a focal point in the Discourse of STEM Education Evaluation.

Theme 3: The influence of STEM education evaluation has multiple dimensions.

Theme 4: Many conversations centered around Program Development for STEM education evaluation.

Theme 1: Presentations Heavily Focused on Learning Experiences in Stem Education Evaluation Contexts

Emerging from the document review of AEA STEM TIG abstracts is the common thread of lessons learned in evaluation theory and field of practice. The basic traits of STEM education academic Discourse emerged when comparing abstract to abstract within a session type (e.g., roundtable or panel discussion), session-type abstracts to session-type abstracts each year (e.g., poster to think tank in 2015), or abstracts each year to abstracts in another year (e.g., ignite in

2015 to ignite in 2016). Many ideas shared by STEM evaluation practitioners showed up as *lessons learned* within the document analysis of the STEM TIG abstracts.

Lessons Learned Defined

In this study, *lessons learned* are experiences drawn from past activities or practices that should be actively considered in future actions and endeavors. Lessons learned are information or understandings that are acquired through experience. The experience could be positive, like a successful STEM faculty transformation program designed to measure impact on student outcomes, or it could be challenging, like evaluating informal STEM programs for young children and their parents. A lesson must be significant in that it has a direct or indirect impact on the evaluation program or project activities. Additionally, it should be applicable in that it identifies a specific design, process, or decision that reduces or eliminates the potential for missteps and setbacks or reinforces a positive outcome or result.

Using the above definition of lessons learned, findings showed that a wide-ranging issue dominated the conversations. Analysis of the abstracts revealed that the lessons learned dealt with a variety of issues and topics. Broadly, some examples of lessons learned might be associated with the Discourse of challenges regarding the recruitment and retention of program participants, implementing large multisite projects, data collection methodology, and collective impact measurement (Table 7). Other lessons learned dealt with professional development matters, including mentorship, preparing for a career in science, and work-life balance. Additionally, the abstract review highlighted that the conversation revolved around the program impact and effectiveness of team-based projects, course activities with STEM program managers, program funding, and program successes.

Specifically, in the academic sector, many STEM evaluators explored the NSF's Broadening Participation agenda while probing Under-Represented Minority (URM) students' experiences with recruitment and retention in the STEM field. More often in the institutions of higher education (IHE), funding agencies like the NSF commission working groups to create frameworks that would address this issue. Through its Research Experience for Undergraduates (REU) Sites program (consisting of a group of ten or more undergraduates who work in the host institution's research programs), the NSF funds many research opportunities for undergraduate students. STEM evaluators then use the created frameworks to guide evaluation processes through stages of implementation, effectiveness, impact, and sustainability (Greene, et al., 2006).

The findings in this study further revealed the lessons learned as described in the implementation, impact, and sustainability of undergraduate research experiences and mentoring components of STEM education programs. In a typical abstract where the presentation evaluated an NSF-funded Broadening Participation project at a Historically Black College and University (HBCU), the discussion hinted at speaking to lessons learned when investigating student affect factors and offered a chance for the audience to learn from their experiences. The abstract highlighted the importance of broadening participation in its opening statement, thus:

> Undergraduate research experiences and mentoring have been identified as promising strategies for broadening the participation of underrepresented minority students in STEM fields. Given the marginalization of underrepresented minorities in science disciplines, often the sense of belonging in STEM and a strong science identity are negatively impacted by the experience of underrepresented minority students.

Table 7. Discussions Surrounding Lessons Learned in STEM Education Evaluation

Sector	Program/Project or Funding/Agency	Evaluation Settings	Program Group/ Community	Dominant Conversation
Academic	NSF REU	PWIs	URMs	Challenges in Recruitment, Retention
	Science and Math Program Improvement	National labs, Clinics, Research Institutes	Students, Teachers, Managers, Evaluators	Project Implementation, Improvement, Collaboration among Stakeholders
	DoE, What Works Clearinghouse (WWC)	Higher Education	Students, STEM faculty	Challenges Implementing QEDs. Impact, and Program effectiveness
Private	Teacher Professional Development in STEM	K–12, Higher Education, Multi-site	Program Partners, Students, and Teachers	Program Implementation and Impact Best Practices, Challenges in PD
	Explore Computer Science Research (eCSR)	Research Institutions, Multi-site	Underserved populations, Women in science	Challenges of large multisite projects, the fidelity of implementation, etc.
Public	National Cancer Institute (NCI)	National labs, Clinics, Research Institutes	Students, Teachers, Project managers and directors	Program Culture, Mentorship, PD, Career Plans, and Work-life Balance
	Junior Reserve Officer Training Corps (JROTC)	JROTC summer academy	JROTC Students	Program Impact, Workforce Trajectories

While a variety of STEM education evaluation abstracts follow a similar route as the empirical case above, others turn their searchlights towards a more theoretical approach in describing their presentations. One such program, funded by NSF Research Training (NRT), described its population sample, STEM graduate students, in the light of Generating, Analyzing, and Understanding Sensory and Sequencing Information (GAUSSI). The discussion summarizes:

> In addition to technical skills, the program aims to support students through tailored/mentored experiences, career counseling support for non-academic STEM career tracks, the creation of a welcoming climate for the majority and minority STEM students, and the enhancement of cross-disciplinary research collaborations.

In the private sector, the abstract analysis showed that programs discussed include teacher professional development in STEM, exploring Computer Science Research (eCSR) in STEM, and projects aimed at moving local communities toward collecting valid evidence for measures used to document festival impacts, among others. Most of the evaluation in an abstract discussion focused primarily on formal educational settings (K–12, higher education, research institute, multisite) or traditional informal educational settings (e.g., museums, science centers, libraries, etc.). However, a handful of the abstracts engaged in conversation around informal education occurring outside the walls of museums and schools. For example, one abstract expanded the conversation to learning lessons through encounters with everyday people in their daily lives and through out-of-home media, social media, and the news media."

Findings showed that evaluators in the private sector work with a wide range of people, including program partners, students, and teachers. In addition, they work with URMs such as women in science, those who are hard of hearing, and people with disabilities. The abstracts talked about program implementation with fidelity and producing outcomes (asking questions

about the type of impact programs have on participants). The discussion also delves into data collection strategies, collective impact measurement, best practices, and challenges in professional development and large multisite projects. The abstracts that focused on informal science education programs described their evaluands as "truly informal settings," while stating the domain of practice and methods as "...subways using out-of-home media, in city parks, and on social media."

The public sector abstracts were able to draw lessons learned from the evaluation of federal agency programs like the National Cancer Institute (NCI) and the US Army Junior Reserve Officer Training Corps (JROTC). The programs evaluated are often situated within national laboratories, clinics, research institutes, and summer academies. Like the academic and private sectors in some respects, the evaluators in this sector work across the board with the school community, including students, teachers, program managers, and project directors. The dominant discussion relates to program culture, climate, and impact.

The public sector conversation leaned toward training, mentoring, and issues of professional development. The extent to which research and evaluation efforts are shaping program participants' scientific preparation, career plans, and work-life balance and experiences play important parts in the Discourse in the public sector.

Theme 2. Professional Development Was a Focal Point in the Discourse of STEM Education Evaluation

(a) The academic sector mentioned funding and shared Broadening Participation (BP) with the private sector (See Figure 4 for the major elements intersecting professional development).

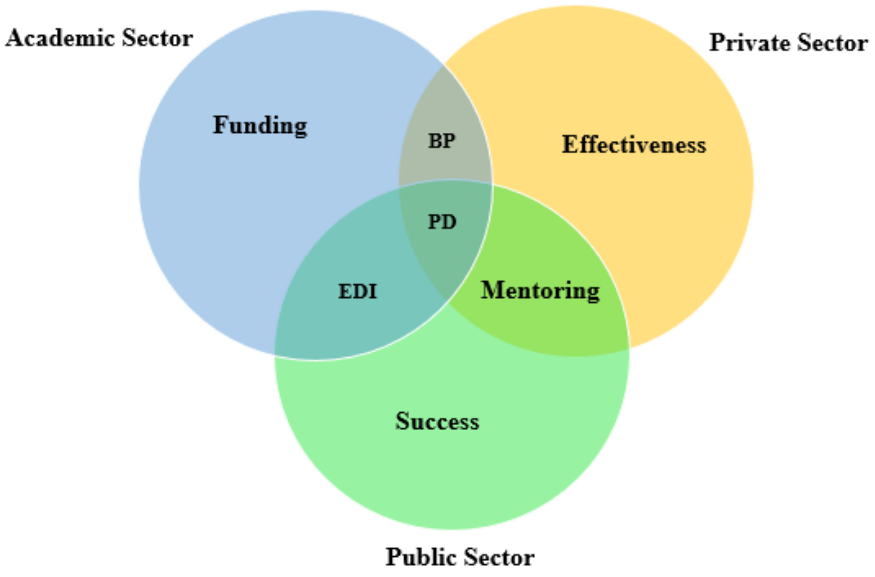
(b) The private sector mentioned program effectiveness most frequently and shared mentoring with the public sector.

(c) The public sector mentioned program successes and shared Equity, Diversity, and Inclusion (EDI) with the academic sector.

Professional Development Defined

Professional Development defined is **as** a set of skill-building processes and activities that are purposefully designed to help specific groups master specific learning objectives. Professional development encompasses the process or processes by which evaluators, researchers, educators, and other practitioners participate to increase their professional competence and broaden their understanding of themselves, their roles, their context, and their careers (Robinson, 2013).

Figure 4. Professional Development in STEM Education Evaluation



Training, workshops, technical assistance, and online or distance learning courses can all be used to provide professional development in evaluation settings. Professional development opportunities can be a powerful tool for networking and gaining a better understanding of STEM educational evaluation projects (Pouezevara, et al., 2014).

Professional Development

The Discourse around professional development efforts in the AEA STEM TIG abstracts revolves around the need to train and improve the skill and competency levels of evaluators and researchers. Abstracts revealed that the discussion paid attention to addressing data collection strategies and challenges in programs designed for increasing participation and performance, providing support, and providing professional development experiences to encourage and inspire URM persistence in the STEM field through to the professoriate level. The discussions in the abstracts also involved challenges and recommendations for designing evaluations for STEM, K–12, higher education, and/or multi-site programs, including the sharing of best practices and effective evaluation designs, methods, and tools among STEM evaluators.

Professional development was mentioned in areas of faculty development programs focused on student-centered course transformation activities and student outcomes. In furtherance of professional development for evaluators, the diversity of experiences in mentorship was highlighted as a part of a continual process of STEM education evaluation. As evaluators shift their attention to and start to commit more to equity issues in evaluation, calls grow for more training, skills, and professional development programs that cater to more inclusive and diverse approaches and methodologies in the STEM field. Findings in the AEA STEM TIG abstracts showed conversations took place around broadening the participation of underrepresented minorities in the STEM field. Findings in this study surfaced a diverse usage of instrumentation that suggest an expansion of existing learning environments in the evaluation field to new learning frontiers. This speaks to the broadening participation agenda espoused by NSF and other funding agencies in ensuring more inclusion, diversity, and equitable access to STEM over time.

In the AEA STEM TIG abstracts, issues related to the professional development of programs, evaluators, researchers, and other practitioners were mentioned by evaluators in all three sectors analyzed: however, to varying degrees. For private and public sector evaluators, professional development conversations show up in after-school STEM programs, among other places. While professional development workshops seem to be the common evaluation setting for all three sectors, academics tend to cater to STEM teachers more frequently, while the private sectors focus on youth, teachers, evaluators, researchers, and designers. Public practitioners pay more attention to the youth, staff, program managers, and administrators.

Specifically, the academic sector evaluators most frequently talked about professional development concerning issues of data collection challenges, training of STEM teachers, and learning opportunities. PD discussed also included designing frameworks, methods, tools, and tips for evaluating K–12 STEM professionals, evaluating the inter-relationship between teacher professional development and student achievement, as well as challenges in transforming their curricula and classroom practice. Additionally, findings show that academic sector evaluators prioritize funding for STEM education evaluation programs that can help support the implementation of effective professional development programs and initiatives. With adequate funding, evaluators, researchers, and educators can collaborate to develop and implement programs that promote equity, diversity, and inclusion in STEM education, and evaluate their impact over time. By continuously assessing and improving upon these programs, they can create a more inclusive and equitable STEM education system that benefits everyone (Table 8).

Dominant conversations in the private sector were about training and development of soft/IT skills for STEM teachers, evaluating and reporting data related to learning outcomes in informal STEM programs, teaching content and facilitating games for youth, how evaluators

come to terms with data that may seem incomplete when evaluating STEM afterschool programs, and how to discuss the data with funders and funding agencies. Professional development plays a critical role in promoting program effectiveness, broadening participation, and mentoring in STEM education. Evaluators may build a more efficient and inclusive STEM education system that benefits everyone, including STEM evaluation practice, by consistently promoting the skills and expertise of STEM educators and other stakeholders (Table 8).

The conversations that dominated the Discourse in the public sector abstracts were the availability of evaluation opportunities for program staff in multiple sites and providing support to improve after-school STEM learning for youth attending publicly funded programs. Others included using a mixed-methods approach in the data collection process and how STEM learning is occurring at a range of out-of-school-time sites. Professional development outreach programs that concern messaging and micro-messaging efforts targeted at making classrooms equitable and increasing student success formed part of the conversation. Professional development plays a critical role in promoting success, mentoring, equity, diversity, and inclusion in STEM education evaluation. A more efficient, inclusive, and equitable STEM education system that serves educators and the public good can be developed by evaluators of STEM education programs by constantly strengthening the knowledge base, building capacity, and broadening the skill set of STEM educators and program organizers (Table 8).

Table 8. Professional Development, Mentoring, Equity, Diversity, and Inclusion; and Broadening Participation

Program/Project or Funding/Agency	Evaluation Settings	Group/ Community	Dominant Conversation
Academic Sector			
Professional Development (PD) Programs	PD Workshops	STEM Teachers	Frameworks, Data Collection, Challenges in PD
Private Sector			
PD: STEAM Teacher Education Programs	Afterschool STEM programs	Youth, Teachers, evaluators, researchers, and Designers	Website Development Training for Teacher, Reporting, Funding Issues
Public Sector			
PD: Afterschool STEM Learning Programs	Publicly funded program workshops in the states	Youths, Staff, Program Managers, Administrators	PD opportunities for Program Staff in multiple sites, mixed methods

Theme 3. The Influence of STEM Education Evaluation Has Multiple Dimensions

Emerging from the document review of AEA STEM TIG abstracts is a discourse on the influence of STEM education evaluation across multiple academic, private, and public sectors. The power to affect evaluation programs, projects, or funding, to affect evaluation settings, or to transform a group of people, or a program community, was highlighted throughout the six-year abstracts analyzed. The discussion in the abstracts included STEM evaluators expressing the influence of using their power of mixed-methods evaluation in exploring case studies of K–12 school-based programs.

The effect of STEM evaluators also surfaced in the evaluation of the African Women in Science Empowerment Model, a mixed-methods evaluation that empowered African women scientists and institutions to deliver innovative gender-responsive Agricultural Research and Development (ARD). Furthermore, STEM evaluators discussed coupling performance measurement systems with program evaluation that is undergirded by a VEE approach. STEM evaluation practitioners produced real-time, interactive data displays of project indicators that helped inform the framing of research questions and the use of the research findings. The findings from that project produced potentially powerful impacts for the Broadening Participation in STEM agenda.

STEM Evaluation Influence Defined

In this study, STEM evaluation influence covers the internal and external processes of the practice, including the role of the evaluator; evaluation as a profession; the ways practitioners use information; the aspect of a program considered; and the purpose, interests, and values inherent in the evaluation findings. Evaluators facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done, from beginning to end, will

affect use. The influence of evaluation concerns how STEM practitioners work and collaborate with programs and participants, funders, and institutions in the real world, apply evaluation findings and experience, and learn from the evaluation process.

In the AEA STEM TIG abstracts, findings (Table 9) showed that the dominant conversation in the Discourse of "Influence of STEM Education Evaluation" was about issues of professional development in areas such as building communities of STEM educators and professionals, building evaluation instruments, administering them, and using them to assess program size and goals. Other topics of conversation revolved around selecting popular media for capturing information, student participation, strategies for communicating with parents in STEM programs and using logic models as effective tools to assist in program planning, implementation, management, evaluation, and reporting ideas for future improvements.

Additionally, challenges and successes encountered in selecting measurable objectives in informal settings and creating outcomes using evaluation frameworks for program implementation and improvement showed up in the discussion. In multisite evaluation settings, public sector evaluators who receive funding from the NSF worked with academic and institutional partners and debated the influence of evaluation concerning accountability for the public good.

The academic sector abstracts discussed several STEM education initiatives reflecting the discursive elements of the influence of evaluation in STEM education programs. For example, the Noyce math and science teacher training programs, the North Carolina State University, as well as PAC (Physics, Astronomy, and Cosmology) projects showed how widespread evaluation influence could go. Evaluation settings discussed range from conferences, college/university, and theoretical and research streams or classrooms. The academics work with conference attendees

such as speakers, exhibitors, and communities of scholars or professionals. They engage kindergarten-through-12th-grade and under-represented high school students. Discussions are centered around building a community of math and science preservice and in-service teachers, using evaluation tools to address issues related to program size, goals, and administration.

Similar to the academics, private sector abstracts delved into a range of contexts demonstrating the influence of evaluation in STEM education programs. For example, two initiatives are informal science education programs and citizen science projects. Evaluation takes place in subways, out-of-home media, city parks, social media, and classrooms. The programming community is made up of town residents, primary school students, and teachers. The challenges and successes encountered in determining measurable outcomes in informal settings are highlighted. Evaluation settings include evaluation frameworks and multi-sites. Program communities include professionals, stakeholders, and institutional partners. The main discussion revolves around program implementation, and improvement, the needs assessment of multiple stakeholders in STEM education settings and developing measurable objectives to fit those needs.

The public sector abstracts mentioned the inclusion of accountability for the public good and offered considerations in support of STEM education program evaluation. Further, the influence of STEM evaluation in the public sector involves conducting needs assessments that are crucial to understanding the specific needs of learners and the programming community. This involves gathering data on students' backgrounds, interests, and learning styles, as well as analyzing the resources and support available in the programming community. The public sector discussion highlights information that can be used to design programs that meet the specific needs of the programming community.

Discussions explored the evaluation of STEM education program implementation, the curriculum, teaching methods, and resources available to support student learning. Examples include the Evaluation Capacity Building Initiative at the Center for Advancement of Informal Science, Technology, Engineering, and Math (CAISE), which was launched to provide connectivity and resources for those designing, implementing, and evaluating informal science, technology, engineering, and math policy learning activities. The Science and Technology Policy Fellowship program sponsored by the American Association for the Advancement of Science is another example that provides opportunities for scientists and engineers to contribute to the federal policymaking process while learning about the intersection of science and policy.

Other discussions in the public sector talked about developing measurable objectives for STEM education programs that involve setting clear, quantifiable goals for student learning, such as improved test scores, increased enrollment in STEM courses, and higher rates of student engagement. Measuring progress against these objectives provides the help needed to assess the effectiveness of STEM programs and inform decisions about program improvement (Table 9)

Table 9. Discussions Surrounding Evaluation Influence in STEM Education Programs Evaluation

Program/ Funding	Evaluation Settings	Group/ Community	Dominant Conversation
Academic Sector NOYCE Math and Science Teacher Training Programs	College/University Theoretical and Research	Scholars or Professionals, K-12, URMs in High Schools	Popular Media, Students and Parent Participation
Private Sector Informal Science Education Citizen Science Projects	Subways: Out-of-Home Media, social media, Classrooms	Citizens of the Community Elementary Students, Teachers	Challenges and Successes in Informal Settings.
Public Sector Heliophysics Education Consortium (HEC)	Evaluation Frameworks	EPO Professionals, Stakeholders	Developing Measurable Objectives, Implementation, and Improvement
NSF Cyberinfrastructure Projects	Multi-site	Academic Institutional Partners	Accountability for the Public Good

Theme 4. Many Conversations around Program Development for STEM Education Evaluation

Program development represented a massive arena of discussion in the abstracts submitted for presentation during the six years at the annual AEA conference included in this study. Findings (**Error! Reference source not found.**) showed a variety of topics dominated the conversation across the three sectors identified. STEM education evaluators addressed issues of programs' professional development geared towards STEM educators and discussed data collection challenges encountered in the process of building instruments and the cost analysis involved.

Other conversations reflected on the steps taken while building survey instruments for measuring a range of outcomes in undergraduate STEM education. The dominant categories of constructs considered include attitudes toward science, scientific behaviors and skills, and cognitive outcomes, such as understanding the process of research and gaining a deeper understanding of the STEM discipline.

A crucial discussion point throughout the program development Discourse was the implementation of professional development for educators that can identify their weaknesses and strengths and track their development. Another finding is that program development supports the evaluation of teacher education programs, assesses the correlation between the level of teacher professional development and student achievement, and addresses issues with evaluation costs and budgets. In addition to building STEM evaluation protocols to address this issue, the use of technology was explored as an alternative facilitator and motivator for students and teachers in learning and advancing a STEM career. The role of external evaluation experts in attending to conditions within a project and in the broader context that can affect project development,

research efforts, and longer-term sustainability and dissemination plans was part of the conversation of program development.

There was a searchlight on data gathering efforts in support of evaluation and challenges in developing a community of practice and common instruments and implementation of these instruments in the informal STEM setting. Other abstracts expanded the discussion on instrumentation and introduced the idea of a searchable repository for informal STEM education (ISE) resources. Powered by the Center for Advancement of Informal Science Education (CAISE), the conversation lit up around providing users with access to project descriptions, evaluation reports and instruments, and research and reference materials. CAISE collects and curates the resources in collaboration with a dozen other professional ISE projects and websites.

Program Development Defined

The basic steps of program development include identifying the problem, planning, implementing, and evaluating the program. Program Development means the ongoing systematic process of planning, implementing, and evaluating a program, which involves a collaborative relationship between evaluators and stakeholders.

Academic sector abstracts take on STEM inquiry in higher education and receive funding from federal agencies (e.g., NSF) and organizations like the Concord Consortium to create large-scale improvements in K–14 teaching and learning through technology. They utilize evaluation probes and models and work in classrooms with K–14 students, URMs, and teachers. The abstracts' discussion focused on issues relating to professional development for teachers, technology utilization in schools, mentoring effect, and program satisfaction.

Private sector abstracts tended to discuss more STEM education training programs and online training frameworks with teachers, with discussion on teacher educational training

dominating some of the conversations. Like academics, the public sector articulated programs like the Community for Advancing Discovery Research in Education (CADRE), and the National Alliance for Partnerships in Equity (NAPE), which are funded through NSF. Additionally, they tend to work across multiple sites, states/districts, or research institutions. The population program community includes doctoral students and URMs (e.g., women and minorities). Dominant in the discussion was professional development for secondary educators; increasing participation, performance, and persistence; and informal mentoring of early career researchers.

Public sector evaluators talked about approaching program development in STEM education evaluation in a holistic manner, with consideration given to each step along the evaluation process. Program planning, implementation, and effective evaluation discussion comprise collaboration among stakeholders, including program administrators, educators, students, parents, and community members who are engaged and supportive of the STEM programs. One typical example in the public sector debate is the National Alliance for Partnerships in Equity (NAPE). The evaluation settings are at either the state or district levels, and program communities are underrepresented populations comprising mostly women and minorities. Major discussion focuses on professional development for secondary educators and increasing participation, performance, and persistence in the STEM education field.

Another representative program is the Community for Advancing Discovery Research in Education (CADRE) network, which is an NSF-funded program. Multisite evaluations are usually the preferred settings or could be a variety of higher institutions with graduate or postgraduate students. Discussion focuses on informal mentoring of early career researchers,

monitoring and measuring the progress and effectiveness of programs, and collecting data on participants' outcomes (Table 10).

Table 10. Discussions Surrounding Program Development in STEM Education Evaluation

Program/Project or Funding/Agency	Evaluation Settings	Group/ Community	Dominant Conversation
Academic Sector Science Inquiry with Technology: the ITSI-SU Project - NSF, Concord Consortium	Evaluation Probes and Models	K-12 students, Teachers	PD for Teachers Technology Utilization
STEM Undergraduate Research	Classrooms	Undergraduate Students	Mentoring Effect and Program Satisfaction
Private Sector STEM Education Training Programs	Online Training Frameworks	Teachers	Teacher Educational Training
Public Sector The National Alliance for Partnerships in Equity (NAPE)	State/Districts	Women and Minorities	PD for Secondary Educators: Increasing Participation
CADRE Network – NSF funded	Multisite	Doctoral students	Informal Mentoring of Early Career Researchers

Question 1a: What are the STEM education evaluation approaches and methodologies utilized in these presentations?

Definitions

The definition of STEM education evaluation approaches covers a broad range of practices or ideas about how to conduct STEM education evaluation. Evaluation scholars and practitioners embrace a variety of terminology that reflects the diversity of expression in the STEM evaluation ecosystem. This study acknowledges STEM evaluator worldview plurality, and an intentional effort is made to accommodate all notions in alignment with the objectives of this dissertation. Examples of alternative words used in the abstracts for approaches to STEM education evaluation include the following: a) models, b) frameworks, c) theories, d) strategies, and e) ways.

STEM Education Evaluation Methodologies refer to a fundamental research strategy that explains the logic or how the research will be carried out and, among other things, identifies the methods that will be utilized in research. It differs from methods in that it does not put a priority on specific methods (the means or modes of data collection). Some examples of expressions used for describing methodologies in the AEA STEM TIG abstracts are (a) modes, (b) processes, (c) ways, and (d) approaches.

Data Used

The data used to answer this question included: (a) document review item—AEA STEM TIG Abstracts

Overarching Finding

A review and analysis of the AEA STEM TIG abstract from 2014 to 2019 were conducted and more than one hundred STEM evaluation approaches and methodologies were

identified. Further review and analysis collapsed the identified approaches into four main themes, each with multiple sub-themes or categories**.** Multiple evaluation methodologies were advanced for the logic behind the choice of methods, and focused coding was used to streamline and sift out the least frequently mentioned ones. The final review included four main themes of evaluation methodologies, each with multiple subthemes. Findings for the evaluation approaches are highlighted first (Table 11 and Table 12).

The Main Themes for STEM Education Evaluation Approaches in the AEA STEM TIG Abstracts

Collaborative Evaluation Approaches are frequently used to describe a range of culturally relevant perspectives, systematically engage, and facilitate co-creation to better understand the complexity of STEM education program evaluation.

Evaluation Capacity Building was frequently mentioned in the abstracts as a learning experience or tool to benefit STEM education program evaluation.

The Systems Evaluation Protocol was used in the STEM TIG abstracts as an underlying philosophy for inquiry and serves as a tool for holistic sense-making of the complexity of STEM education program evaluation.

Evaluation Frameworks for STEM Education are often used interchangeably as models; frameworks were the most frequently mentioned approaches to STEM education evaluation in the AEA STEM TIG abstracts.

STEM evaluators have a plethora of evaluation approaches available to them. They address complex challenges and contemplate evaluation developments for current and future use.

Collaborative Evaluation Approaches are widely used in the abstracts to describe the range of culturally relevant perspectives in program communities, systematically engage

stakeholders or partners, and facilitate the co-creation of a more complex understanding of programs (Table 11).

Table 11. Typical Evaluation Approaches in the AEA STEM TIG Abstracts, 2014–2019

No.	Collaborative Approaches (CA)	Evaluation Capacity Building (ECB)	Systems Evaluation Protocol (SEP)	Evaluation Frameworks (EF)
1	Culturally Responsive	Repository	Systems Thinking	Logic Model
2	Values-Engaged Educative	Empowerment	Evolutionary	Theory of Change
3	Social Network Analysis	Exemplary	Evaluative Thinking	Rasch Model
4	Utilization-Focused	Technology-Based	Online cyberinfrastructure	Pedagogical
5	Equity-Focused	Best Practices	Systems Thinking Orientation	Transactional model
6	Developmental	Multisite		Structural Equation Model
7				Pathways

Table 12. The Dominant Evaluation Approaches in the AEA STEM TIG Abstracts by Evaluator Sector

Dominant Approaches	Academic Sector	Private Sector	Public Sector
Collaborative	Develops Instrumentations, etc.	Same as Academics	Impact of Education Interventions
Formative/Summative	Support STEM Undergrad Women	Formative Feedbacks	Communication with Funders
Utilization-Focused Evaluation (UFE)	Building communities	PD: STEM out-of-school	Emerging Technologies
Values-Engaged Educative (VEE)	Evaluation Working Groups Responsive Evaluation Methods		Utilize VEE to Address Public Demands for Accountability
Culturally Responsive Evaluation (CRE)	Stakeholders from NSF share learnings and best practices	Museum of Science and Industry	Discuss Factors related to Alliance Projects
Systems Thinking, etc.	Communicate with key Stakeholders	Learn Spatial Concepts	Institutional Change, Cultural shifts
Logic Model (LM) / Theory of Change (ToC)	Guide Decisions: Addresses Challenges Program Challenges	Framework for Action and Accountability	Academy Experiences, STEM, or non-STEM academic and workforce
Developmental Evaluation (DE)	Situate STEM Evaluation Activities in the Context of Student Growth		Diverse population of scientists and clinicians into sustainable careers
Social Network Analysis	Provide Program Actionable insights		Clarity in Program Social Networks and Perceptions

Evaluation Capacity Building was highly mentioned in the abstracts as a learning experience or tool to benefit STEM education program evaluation (Table 11.)

The Systems Evaluation Protocol was used in the STEM TIG abstracts as an underlying philosophy for inquiry and serves as a tool for holistic sense-making of the complexity of STEM education program evaluation.

Evaluation Frameworks for STEM Education are often used interchangeably as models; frameworks were the most frequently mentioned approaches to STEM education evaluation in the AEA STEM TIG abstracts. The frameworks are generally designed to provide actionable objectives and accountability to strengthen and improve STEM education programs.

Additionally, the frameworks were mentioned as adaptable evaluation tools built to measure outcomes, impact, and progress rather than operating within an overarching strategy that sets out clear definitions and expectations. Underlisted is examples of frameworks or models used by evaluation practitioners while writing abstracts for presentation at the AEA conference. Extracts from the abstracts are included where appropriate and for illustration purposes (Tables 11 and 12).

The Main Themes for STEM Education Evaluation Methodologies

This prescribes the choices made by evaluators in the process of gathering data and explains why they were made. Justifications for study design choices are surfaced in research methodologies through the explanation of the chosen methods being the best fit for the research aims and objectives and would offer valid and trustworthy findings. The findings highlighted below to reveal the main themes for STEM education evaluation methodologies.

The main Evaluation Methodologies identified in the AEA STEM TIG Abstract review:

Traditional Methodologies are characterized by representations of experiences and sense-making that adhere to industry-validated and standardized ways of knowing and being (i.e., epistemological, and ontological assumptions, respectively). Findings from many of the abstracts analyzed showed STEM evaluators expressed their methodologies through these traditional epistemological and ontological formations. These STEM evaluation practitioners connect the work they do to the methodologies they embrace by explicating it through inquiries that are well-known or established in the past. Listed below are major findings from the abstracts that highlight the traditional evaluation methodologies (Table 13 and Table 14).

The Zigzag Methodologies are opposed to normative components of sense-making and pursue non-linear, meandering paths and connect the articulation of findings (a process of truth-telling) with theoretical orientations that drive the choice of methodologies employed in evaluation. Underlisted is some examples of categories identified in the abstracts that used the Zigzag methodologies in the choice of their methods for evaluation activities (Table 13 and Table 14).

- Research exploration in science and mathematics educators' experiences.
- The academic coaching process and experience Intersectionality for Indigenous evaluation efforts
    - Unpacked and explained using real-world examples to highlight some of the diversity within and across Tribal/First Nations communities.
    - In theory, method, design, and implementation: it is possible to create evaluations that are responsive to culture and community but also are scientifically rigorous.

Mixed-Methodologies entail the philosophical assumptions underlying the application of qualitative and quantitative methodologies, and the incorporation of both approaches in tandem so that the overall strength of a study is greater than either of them (Creswell & Plano Clark, 2017). The most frequently used methodologies that used multiple data sources as the preferred method to make data-based decisions and report on findings regarding STEM education evaluation are listed in Table 13.

Scientific Methodologies establish assumptions about the regularity of causation and focus interrogation of inquiry as to what (happened) rather than to the why (justification) it happened. Rather than emphasize the importance of understanding lived experiences and a plurality of worldviews, scientific methodologies are characterized by procedures and methods designed to prioritize and privilege general laws with a view on reproducible facts and outcomes. Listed in Table 14 are some of the scientific methodologies that showed up in the STEM TIG abstracts for presentation at the AEA.

Table 13. The Typical Evaluation Methodologies in the AEA STEM TIG Abstracts

Traditional Methodologies	The Zigzag Methodologies	Mixed-Methodologies (Quant. and Qual.)	Scientific Methodologies
Survey	Research exploration	Quantitative Biology Concept Inventory (QBCI)	Experimental methods ● Randomized Controlled Trials (RCTs)
Interviews	Academic coaching process and experience	Statistical Techniques	Nonexperimental methods ● Quasi-experimental Designs (QEDs)
Document/ Systematic Review	Intersectionality for Indigenous evaluation	Examination of Program Records	
Observation			
Focus Group			
Case Study Design			
Videos/Journals			
Curriculum Assessment			

Table 14. The Dominant Evaluation Methodologies in the AEA STEM TIG Abstracts by Evaluator Sector

Methodologies	Academic Sector	Private Sector	Public Sector
Traditional Methodologies	Observation– Students (college/university setting). Validated Student, Teacher, and Principal Survey Instruments	Interviews–Researchers and Practitioners (CAISE)	Surveys – Students from Academic Departments nationwide (CERP program)
	Reflection and Insights – Medical Students (Online Learning Initiative)	Scales Development–(Citizen Science Projects)	Case Study – STEM faculty (NSF ADVANCE Program)
Mixed Methodologies	Surveys– Students, Teachers, and Principal; Focus Group–Teacher and Student Interviews–Grant coordinators.	Surveys, observation, and focus groups–Middle school students (Informal STEM Education)	Quasi-Experimental Designs Surveys, Interviews, and Focus Groups–Parent and Team Leaders
Scientific Methodologies	Quasi-Experimental Design, Matched Classes–various STEM Disciplines		Randomized Control Trials, Quasi-Experimental Design, Regression Discontinuity Design,
Zigzag Methodologies	The Learning by Making Model in High Schools with Students	Combination of Theory, Method, Design, and Implementation–Responsive Indigenous Eval.	The Expert Application of Theory and Methods–(Alberta Innovates Health Solutions: AIHS)

Question 1b: What are the differences between evaluators who work in academic institutions versus practitioners in private and public sectors?

Data Used

Data used to answer this question included: a document review of AEA STEM TIG abstracts, 2014–2019.

In answering question 1b, abstract data were disaggregated to contrast academic, private, and public sector STEM evaluators along several significant components, including evaluation settings, group/community, and dominant issues that showed up in conversation. The findings begin with highlights of the differences between academic, private, and public sector evaluation settings and the major similarities or points of convergence. Next is a description of the characteristics of the key evaluation setting and program participants in terms of the group or community with which the STEM evaluators work—for example, how the evaluators describe the evaluation setting or community of students in elementary, middle, and high schools; colleges; and universities in the academic, private, and public sectors. Following that is a comparison of the main items that dominate the discussions of the abstracts of academic, private, and public sector evaluators. The final section presents findings for the varying approaches and multimodal methodologies that are used by academic, private, and public sector evaluators in different contexts of STEM education evaluation.

Overarching Finding

The research used a document review and content analysis as a guiding framework to meaningfully engage AEA STEM TIG abstracts. A summary of the study's findings, including major insights from the abstract analysis, is provided below. As previously stated, the findings are categorized into four contexts, including (a) STEM Education Programs,

(b) Evaluation Settings, (c) Population Group/Community, and (d) Dominant Conversation. These four contexts are then used to surface the differences between evaluators who work in academic institutions and practitioners in the private and public sectors. The highlight of key findings includes the following:

STEM Education Programs

Academic sector evaluators most frequently discuss research experiences for undergraduate programs. The evaluation settings range from K–12 classrooms, STEM learning, and research centers, and summer institutes, primarily white institutions (PWI) engineering schools, Historically Black Colleges, and Universities (HBCUs), and Hispanic Serving Institutions (HSIs). Academic STEM evaluators work with predominantly K–12 students, teachers, college students (undergraduate, and graduate), STEM faculty, and underrepresented minority students (URMs) from across the nation.

Topics dominating the conversations among evaluators in the academic sector are current and future program planning and assessment of STEM initiatives with evidence-based theory, and the implementation of such efforts. Further, they discussed issues of stakeholder groups and utilizing multiple data sources, including technology and big data, and shed light on the difficulties and triumphs that we encountered in practice. They paid attention to historical and institutional dimensions of culture that informed the evaluative approach and analyses of their practice. They mentioned different theories, such as resiliency theory, peer-assisted learning theories, social capital theory, intersectionality, and cultural capital theories that shaped their work. Undergraduate research mentorship, building STEM communities, using evaluation instruments, and implications for practice were part of the discussions.

Lastly, the process of selecting popular media, how to get students to participate and meaningfully engage them in STEM, parental involvement and communication, professional development for teachers, and use of technology in practice were also part of the dialogue (Table 15).

Numerous distinct programs were discussed by professionals in the private and public sectors. While they share some similarities, the differences stand out with private sectors engaging in conversations around "Informal Science Projects," and "Out-of-School-Time STEM Programs," whereas, the public practitioners talk more about the Inclusion across the Nation of Communities of Learners of Underrepresented Discoverers in Engineering and Science (NSF INCLUDES), and the National Aeronautics and Space Administration (NASA) Office of Education programs. In addition, public sector evaluators talked about programs that recognize underrepresented minority (URM) students and professional development experiences that help persist students in the STEM pipeline through to the professoriate (e.g., NSF AGEP).

Evaluation Settings

Evaluators from all three areas, Academic, Private, and Public Sector, frequently mentioned multi-site (e.g., higher education, research institutions/centers) as a domain of practice. Academic and Private sector evaluators are synonymous with K–12 schools and student classrooms. In terms of the evaluation environment that receives the greatest attention, colleges and universities distinguish academics from evaluators in the private and public sectors. Furthermore, museums and community learning centers, subways, out-of-home media, city parks, and social media were most frequently mentioned by private sector practitioners. Academic partnerships, clinics, and state or district levels are most responsible for STEM evaluations in the public sector (Table 15.)

Population Group/Community

Academic and Private sectors are similar in the discussion about K–12 students, teachers, undergraduate and graduate students, faculty, and underrepresented minorities. All three sectors—academic, private, and public—discussed underserved populations, women in science, designers, other evaluators, researchers, program partners, and other stakeholders in program communities. All three sectors—academic, private, and public—frequently mentioned program organizers, managers, partners, and communities of scholars or professionals in their abstracts for presentation. Private sector evaluators are differentiated in how they communicate with citizens or members of the community at large or in and around the nation.

Dominant Conversation

The Discourse of STEM Education Evaluation in the AEA STEM TIG abstracts varied widely. However, some commonalities existed. In all three sectors, evaluators engaged in conversation around program planning and implementation, challenges and successes, impact on the program community, improved classroom implementation, and student successes. There were conversations around issues of student participation and parent communication, cultural context, different theories, undergraduate research mentorship, building STEM communities, and using evaluation instruments that were most frequently discussed by both academic and private sectors. All three sectors coalesced around the topic of professional development for educators and the use of technology in practice. Additionally, challenges and successes of large multisite projects and data collection strategies in practice were common items of discussion among all three sectors of academic, private, and public evaluators. Educational training and new thinking about program designs, best practices, creating and using measures, and organizational practice were found in private sector conversation and not lacking in academics as well.

The topics of discussion that differentiated each sector were few and included a thin marker for private and public sectors, respectively. These include the following: (a) selecting popular media was synonymous with private sectors, and (b) regarding commitment towards accountability for the public good, public-sector evaluators stood apart when compared to those from academic and private sectors.

The academic, private, or public sector STEM education evaluators engage in the evaluation of programs or projects in different evaluation settings and work with populations, groups, or communities while engaging in a wide range of discussions and addressing many topical issues. The dominant Discourses of STEM Education Evaluation in the AEA STEM TIG evaluation abstracts have been identified earlier in this study and will be used in the presentation of findings in this section.

Comparison of STEM Evaluators in Academic, Private, and Public Sectors

The findings of the similarities and differences between evaluators who work in academia versus practitioners in private and public sectors are summarized in this section (Table 15).

Academic Compared with Private and Public Sector Evaluators

When comparing academic, private, and public sector STEM evaluators, it is essential to consider their areas of focus and the populations they work with. Academic evaluators work across various settings, including STEM learning and research centers, K–12 school classrooms, and colleges/universities. They collaborate with different groups or populations, such as undergraduate and graduate students, program organizers, managers, teachers, URMs, and other communities of scholars or professionals.

The implications of comparing academics with private and public sector STEM evaluators are significant, particularly in understanding how they approach STEM education

evaluation. For example, the findings from the content analysis of the AEA STEM TIG abstracts show that academic evaluators tend to focus on program challenges and successes, impact on the program community, improved classroom implementation, and student success. They also consider issues of cultural context, different theories, undergraduate research mentorship, building a community of Math and Science service teachers, and using evaluation instruments for gathering and analyzing data.

Moreover, academic evaluators are invested in the conversation around selecting popular media, student participation, parent communication, professional development for teachers, and the use of technology in practice. These areas of focus are important for promoting equity, inclusion, and diversity in STEM education, particularly among underrepresented minority groups.

In contrast, private and public sector STEM evaluators may have different areas of focus, depending on their organizational goals and funding sources. Private sector evaluators may prioritize program impact on revenue, customer satisfaction, or innovation, while public sector evaluators may prioritize program impact on policy or social outcomes.

Overall, understanding the different areas of focus and populations served by academic, private, and public sector STEM evaluators is crucial for ensuring the effectiveness and sustainability of STEM education programs. By leveraging their unique perspectives and expertise, evaluators from various sectors can work together to promote equity, inclusion, and diversity in STEM education and create a better future for all learners.

Private Compared with Academic and Public Sector Evaluators

The comparison between private, academic, and public sector STEM evaluators is important because it sheds light on the similarities and differences in their focus areas, priorities, and practices. This knowledge can inform decision-making and improve the effectiveness of STEM programs across sectors.

The findings suggest that private sector evaluators are comparable to academic sector evaluators in terms of the evaluation settings they work in and the populations they serve. They both work with K-12 classrooms, higher education, research institutions, and multi-site programs, as well as underserved populations, women in science, designers, and other stakeholders. This implies that private sector evaluators have the necessary expertise and experience to contribute to STEM evaluation efforts alongside academic evaluators.

Moreover, the fact that private sector evaluators work in museums and science centers, community learning centers, subways, out-of-home media, city parks, and social media suggests that they have a broader reach than academic evaluators who typically work within academic institutions. This implies that private sector evaluators may be better positioned to evaluate informal STEM learning experiences and programs that take place outside of formal academic settings.

The finding that private sector evaluators place a premium on best practices, creating and using measures, and organizational practice suggests that they may be more focused on practical aspects of evaluation, such as fidelity of implementation and data collection, than academic evaluators who may be more concerned with theoretical frameworks and research questions. This implies that private sector evaluators may be better positioned to provide actionable recommendations for improving STEM programs and practices.

Overall, the comparison between private, academic, and public sector STEM evaluators highlights the strengths and expertise of each sector and emphasizes the importance of collaboration across sectors to improve the effectiveness of STEM programs and initiatives.

Public Compared with Academic and Private Sector Evaluators

The comparison between public, academic, and private sector STEM evaluators have several implications that can shed light on the strengths and differences of these professionals. The findings suggest that while there are similarities in terms of the types of evaluation settings, student populations, and evaluation frameworks, the public sector evaluators stand out in their commitment to accountability for the public good.

Public sector evaluators tend to work with a variety of research institutions, summer academies, seminars, and workshops, which aligns with their academic and private sector counterparts. However, they prefer to evaluate post-secondary academic alliances, national laboratories, clinics, and state/district-level work rather than classroom settings. This could be due to the focus on program implementation, improvement, impact, assessment, professional development, mentorship, scientific preparation, and developing measurable objectives in their work.

In terms of the student population, public sector evaluators work with undergraduate and doctoral students and underrepresented minorities, including women and minorities, which is in line with academic and private sector evaluators. They also work with teachers, project directors, professionals, stakeholders, academic institutional partners, evaluation practitioners, and stakeholders in programs or initiatives.

One significant difference between public sector evaluators and their academic and private sector counterparts is their emphasis on accountability for the public good. Public sector

evaluators are committed to improving STEM education for the benefit of the public and society. This may manifest in a focus on program culture and environment, career plans, work-life balance, and workforce trajectories. Additionally, public sector evaluators tend to prioritize multi-scale evaluations, impact frameworks, and culturally responsive and inclusive evaluations, in collaboration with stakeholders.

In conclusion, understanding the implications of comparing public, academic, and private sector STEM evaluators can help to identify the strengths and differences of these professionals. The commitment to accountability for the public good is a unique aspect of public sector evaluators that distinguishes them from their counterparts. This emphasis on improving STEM education for the benefit of society can have far-reaching impacts on the future of STEM fields and the workforce.

Table 15. Differences Between Evaluators in Academic Versus Practitioners in Private and Public Sectors

Context	Academic	Private	Public
STEM Education Programs	K12 STEM Education, Experience for Undergraduates, K12 Teacher STEM	Informal Science Projects, Out-of-School Time STEM Programs	The NSF INCLUDES Program, NASA Science Research
Evaluation Settings	K–12 Schools, Colleges/Universities, STEM Learning, and Research Centers	Same as Academic: Community Learning Centers. Social Media.	Research Institutions, Summer Academies, Seminars, Workshops
Group/Community	K–12 Students, Teachers, College Students, Faculty, URMs, Stakeholders	K-12 Students, and Teachers Women in Science, Citizens	College Students, and URMs, Institutional Partners, Practitioner
Dominant Conversation	Program Planning and Implementation, Challenges and Successes, PD, Impacts	Same as Academics, Collective Impact Measurements in the Informal Evaluation Settings	Same as Academics, Mentorship, Commitment to Accountability for the Public Good

Question 2: What are STEM evaluators' thoughts and impressions about the current and future state of the field?

Definitions

Thoughts and impressions of the current and future state of the STEM education evaluation field are defined as follows:

Thoughts: Conceptions or opinions produced by thinking, or occurring as a result of reflection on past, present, or future events.

Impressions: a viewpoint, idea, feeling, or belief about the STEM education evaluation landscape formed based on surface evidence.

Current State: How STEM education evaluators are describing the process of their work right now. In addition, how they are describing the challenges they face with that process and all the other steps and actions involved.

Future state: The greater vision for the STEM education evaluation field. It is the place where evaluators want the field to be considering the evolving state of technology and how that applies to their present evaluation approaches and methodologies.

Equity: The disparities that exist between population groups. It is based on the notions of fairness and social justice. The disparities that exist between population groups that are avoidable and unfair are termed inequities. STEM education program evaluators who embrace the Social Justice paradigm seek to respond to the Broadening Participation issues (equity, diversity and inclusion, access, etc.) when developing and implementing an evaluation to close the gaps.

Levers of Equity: Means of exerting pressure to accomplish the strategic goal of equity.

Overview of Interview Participants and Use of Pseudonyms

In reporting the interview data, I used pseudonyms to balance the confidentiality of the evaluators who graciously agreed to be interviewed and to humanize their voices. During the interview phase, it was discovered that some evaluation practitioners were concerned about confidentiality and were hesitant to share their experiences and perspectives openly. I promised them that I would keep specific information confidential and would not report anything that would link them directly to the data. I had to remind interviewees that I am bound by the ethics of the IRB informed consent for human participants.

My memos and reflections revealed that taking this approach allowed the evaluators I interviewed to feel more at ease and confident in speaking candidly, resulting in more valuable and insightful information being gathered. As a result, the following Table 16 provides brief background information for interview participants. The findings are organized by themes, and where appropriate, supporting quotes were inserted to demonstrate the contribution of the 13 participants in this study.

Data Used

The data used to answer this question included: (a) STEM evaluator interviews, and (b) document review items—AEA STEM TIG evaluation abstracts.

Overarching Finding

The researcher used the co-constructed interview protocol as a guide to meaningfully. engage STEM education evaluators in semi-structured interviews. The following themes emerged during data analysis and are presented below in support of the overarching finding:

The Key Issues for STEM Evaluators' Perspectives on the Field's Present and Future.

Theme 1. Lessons Learned about Broadening Participation with a focus on equity were central to the Discourse of STEM Education Evaluation.

Theme 2. STEM education evaluators had positive things to say about professional development initiatives and community organizing practices.

Theme 3. The era of accountability has evolved, as seen through the lenses of leadership, social justice, and self-socialization.

Theme 4. Data Utilization in the Age of New Technology is significantly influencing how STEM education programs are evaluated today and for the future.

Theme 5. Discussions on approaches and methodologies are leading to the next frontier in the evaluation of STEM education programs.

Table 16. Interviewed STEM Evaluators' Backgrounds and Pseudonyms (STEM Evaluation Practitioner in Academic, n = 6; Private, n = 5; Academic/Private, n = 1; Public, n = 1)

Pseudonym	AEA STEM TIG linkage	Evaluation Approaches and Methodologies	Sector: Years of Experience
Sade	Yes	Collaborative Approaches (CA), Feminist Theory (FT), Culturally Responsive Evaluation (CRE)	Academic: 16–20
Tshepiso	Yes	CA, Quantitative and Qualitative Methods (QQMs), Educational Statistics (EDs.)	Academic: 6–10
Eziamaka	No	CA: Values-Engaged Educative (VEE), CRE, Social Justice (SJ) Oriented. Quantitative Methodologies (QNMs), Mixed Methodologies (MMs)	Academic: 16–20
Mpho	No	CA, Community of Practice (CoP), Qualitative Methodologies (QLMs)	Academic: 11–15
Eghedo	No	CA, CRE, FT, MMs	Academic: 16–20
Vimbai	No	CA, Zigzag Methodologies (ZZMs)	Academic: 0–5
Adaora	Yes.	Systems Thinking Orientation (STO), CA, ZZMs	Private: 20-25
Latifah	No	CA, CRE, FT, Allyship, MMs	Private: 11–15
Ugochukwu	Yes	Empowerment Evaluation, ZZMs	Private: 16–20
Adekunle	No	CA, CRE & Equitable Evaluation, MMs	Private: 16–20
Tlamelo	Yes	CA, FT, CRE, Community Engagement, MMs	Private: 6–10
Ofentse	Yes	CA, QQMs, ES	Public: 11–15
Rolake	Yes	CA: VEE, CRE, SJ Oriented. QNMs, MMs	Academic/Private: 16–20

During the interviews, STEM evaluators in dialogue with the researcher more often describe personal or professional experiences with their approaches. The findings in this study identify and explain the different value systems among participants, which undergird and illuminate their perspectives in the STEM education evaluation ecosystem. Essentially, STEM evaluators bring with them different values, are trained differently, have divergent interests, and have an emerging and evolving paradigm regarding approaches to STEM education evaluation. In addition, in their line of work, interviewees articulated numerous and sometimes distinct schools of thought (Carter & Little, 2007) for conducting STEM education evaluation.

This array of reconstructed logic that lends justification to choices of methods used in theory and practice is termed methodologies. This study is not an exhaustive list of evaluation methodologies; however, through dialogues with STEM evaluation practitioners, the researcher detailed how the Discourse of STEM Education Evaluation is taking place in domains of practice. The majority of the distinct methodologies articulated by STEM evaluators in this investigation are driven by (a) the type of questions asked by the program/project or evaluation undertaking, (b) the needs of specific program organizations, (c) the values espoused by the program owners or funders, and (d) values of evaluators and consideration for the programming community.

The most cited methodologies emerging from interview discussions with STEM evaluators are mixed methodologies, followed by collaborative methodologies. Case studies and survey methodologies are most prevalent in the Discourse of STEM Education Evaluation where evaluators pay attention to issues of identity, status, and power in the domain of programming. In dialogue with evaluators, the exploration of identities and voices are key elements in both the process and the content of the exchange.

Sample

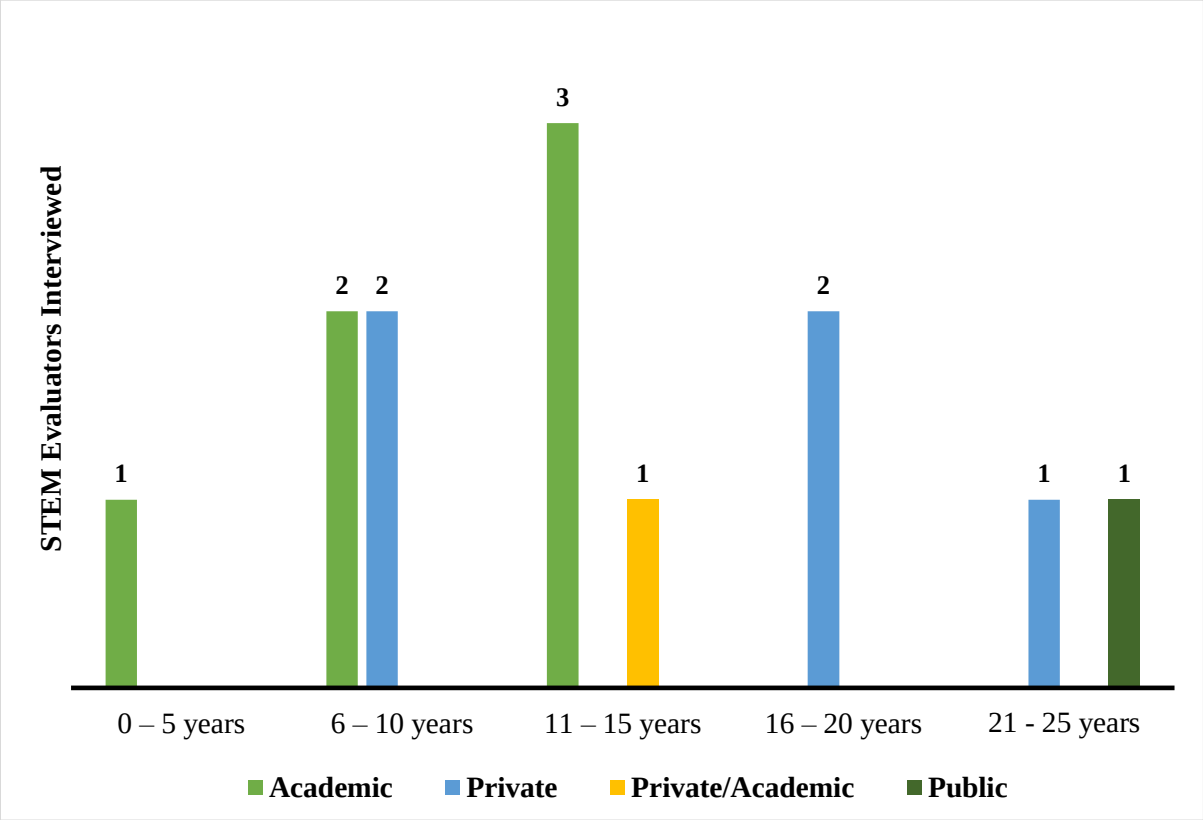
Thirteen participants (N = 13) were interviewed for this study, and their demographic data is presented in Table E19 of Appendix E. The table shows that five (38%) of the participants were African American women, with three from the academic, one from the private, and one from the private/academic sector. There were four (31%) White women, with one from the academic, and three from the private sector. One (8%) African American woman identify with the private/academic sector and one (8%) Hispanic/Latino woman is represented from the academic sector. Three men participated in the interviews, one African American (8%) from the academic, one White (8%) from the private and one Asian (8%) from the public sector.

Many of the interviewees did not start their professional career as STEM evaluators, they mostly "stumbled into the field of evaluation" as many others in the profession. However, all six academic sector participants earned a Ph.D. with 3 of them having a concentration in program evaluation. Four of the 5 privates and 1 participant in the public earned a doctorate with no concentration in evaluation, and the private/academic sector participant earned a Ph.D. with a concentration in evaluation.

Participants' academic backgrounds range from economics, sociology, psychology, systems thinking, policy studies, sciences, business, or a related field. They started their professional career trajectory as high school classroom teachers, researchers and evaluators in public education departments and private evaluation firms, graduate and teaching assistants, and faculty members in academic institutions. The total years spent in STEM education evaluation varied among the 13 participants. One participant had 0-5 years of experience. Those participants with 5–10 years and 10–15 years of experience represented 31% of the sample each. Those

participants with 15–20 years and 20–25 years of experience represented 15% of the sample each. (See Figure 5 below, and Table E20 in the Appendix E).

Figure 5. STEM Evaluator Interview Participants: Years in Evaluation Practice



This section presents the accounts put forward by the thirteen STEM evaluators and explains their perspectives regarding the current and future state of the STEM evaluation field. Specific attention is paid to how STEM evaluators characterize Discourses connected to evaluation approaches as defined in this dissertation as a broad range of practices or ideas about how to conduct STEM education evaluation, including a plurality of worldviews that align with the basic tenet of the Discourse in this study. Additionally, the study used semi-structured interviews to better understand the STEM evaluator's perception of methodologies which justify the choice of methods used in their theory and practice.

As earlier stated, five major themes emerged from the analysis of interview data regarding STEM evaluators' thoughts and impressions about the current and future state of the field. Summarily, they include (a) Discourse on Broadening with equity as the main focus, (b) Successes and Challenges experienced in STEM evaluation programming, (c) a New Era of Accountability through the prism of leadership, social justice, and self-socialization, (d) Technology and Big Data taking over and reshaping the field of STEM education program evaluation, and (e) Discourse of Approaches and Methodologies moving the STEM education evaluation landscape to the next frontier. Even though evaluators' thoughts and impressions have been distilled into the themes, there are several overlaps and more so interconnecting strands of elements showing up in multiple categories and creating a complex web of Discursive properties.

Theme 1: Lessons Learned about Broadening Participation with a focus on equity was central to the Discourse of STEM Education Evaluation.

The following sub-themes have been identified in the analysis of interview data and used in the presentation of findings for the theme.

A. Broadening Participation with a focus on equity (with Diversity and Inclusion)
B. Inclusion of Voice and Power (diversity of programs, participants, and population groups)
C. Attending to STEM education evaluation by embracing: (a) Evaluator Values, (b) Program Effectiveness and Improvement, (c) Program Impacts

A. Broadening Participation with a Focus on Equity

Lessons learned is an umbrella term used in the Discourse by interview participants to describe their Broadening Participation experiences during the planning and implementation of programs in a variety of settings, including multisite, higher education, K–12, science centers, and museums when working with leadership teams on funded initiatives. In this study, evaluators

engage in discussions that show they are more aware of the range of evaluation program needs and cognizant that not one size fits all. They are also mindful of what the realities of the programming community look and feel like in the context in which the different initiatives are taking place.

Findings suggest that STEM evaluators position themselves to understand the program participants, what their needs are, and how the project may be serving them. They work closely with the leadership team, learning to understand their needs and how the projects may impact all the different stakeholders or partners within the domain of practice. They ask questions at every stage and scrutinize every assumption in their quest to understand what works and why. In the inquiry, they search for the truth behind equity, diversity, and inclusion; they encounter challenges, some of which are age-old and institutionalized, and attempt to uncover the truths behind the stereotypes and the generalizations. Specifically, analysis of the interview data reveals the following:

One hundred percent of academic, academic/private, and public sector participants, and two-fifths (40%) of the private sector participants mentioned lessons learned when discussing the issues of EDI under the umbrella of Broadening Participation. Four evaluators, one from each sector of the interviewees, captured the essence of what the participants shared when asked about their thoughts and impressions about the current and future state of the field. Vimbai gave her opinions and reflected on her own experiences to give various types of scaffolding for the programming community and leadership teams they have been working with:

> So, it is learning how to better serve these populations. And the idea is that the evaluation team can serve as a learning platform, let us say for the project team, so they can learn about the implementation and how to improve what is happening and just remember that

all the equity and diversity and, you know, all of these issues should be at the front. You know, like, that should be like, and how to say that everyone should be very aware of all these issues. – Vimbai, Academic Sector Evaluator

Another participant shared the multi-faceted aspects of what she believes EDI should look like for the current and future state of the STEM evaluation field. She expanded her thought by emphasizing the need to turn the searchlight beyond formal to informal settings and providing more opportunities to listen to alternative voices and restructure the field in service of more inclusiveness of the underserved population.

…well, so I think, and I hope that there's going to be more informal learning. So, in informal learning, even pre-pandemic, even pre-George Floyd, important voices were saying, "Hey, like, there is such a divide in who participates in informal science." And it's not just museums, and it's not just science clubs, and it's not just citizen science. And it's like everything, all informal, like every sector of it, is most people who are in who are fluent and highly educated, and often White. And there were voices kind of saying, "Cut it out!" years ago. And a lot of us didn't hear that as loudly as we should have. And now people hear it and can think about structuring the work differently. – Adaora, Private Sector Evaluator

Another evaluator who is at the nexus of private and academic spaces says their EDI experiences working in both academic institutions and private firms have taught them lessons of inclusion. Working in both sectors, Rolake, the evaluator, has learned to value the idea of inclusion when networking with colleagues and talking more about how they are thinking about naming the people they work with. The goal of the nomenclature to keep in mind during interactions among evaluation colleagues is to disambiguate the spoken or written names of

people involved in the evaluation process. Coming up with these distinctions is seen as critical to the identity of programs and the overall EDI agenda. Rolake commented further:

> ...I feel like stakeholders don't do all the time for you, stakeholders, placeholder, but it doesn't identify who we're talking about. So, we are working towards saying, okay, for me, if I mean, someone that's part of the organization, then naming that, as you know, station or something like that, it was a community partner, calling them a community partner. So, it's both being respectful if it is not a statement that someone chooses to be identified with. But also, being more specific about who we mean, who we're talking about, when we are working with clients or community members, that makes sense?
>
> Rolake, Private/Academic Sector Evaluator

For evaluators working in the public sector, lessons learned in using technology to better inform our practice and move the conversation around EDI forward are paramount to the Discourse of STEM Education Evaluation. This evaluator in the public sector wants to see more conversations with more people and is interested in knowing how this plays out in their practice. The evaluator joined and became actively involved in a technology team to immerse himself in the ongoing conversation about the use of technology to better serve evaluation clients. As the conversation continues in the technology space, more and more questions are emerging about decision-making improving STEM evaluation practices. The complex issues in STEM education programs require expanding the conceptions of EDI to evaluate engagement in a close relationship with technology use. Ofentse, the evaluator commented below:

> ... so, the more we grapple with these questions, and that is, the more that improves our practice. And then we can make good decisions in terms of our practice. So, I feel like I know how I use technology, in my practice, I think is to the benefit of my clients, not

necessarily, you know, and to me and the way we connect, you know, how it's formed our relationships. – Ofentse, Public Sector Evaluator

B. Inclusion of Voice and Power

Analysis of the interview data showed that STEM evaluators described using their voices and power to elevate the conversation in the practice and service of the programming community. In reflecting on lessons learned during dialogue with the researcher, STEM evaluators expressed the challenge and concern that diversity of participation is often limited to prominent or articulate program stakeholders, notwithstanding pertinent issues affecting minority groups that may be present and need equal, if not more urgent, attention. To address this gap, evaluators tend to find creative ways to include the voices of groups that are affected by a program using unique evaluation approaches, a variety of skill sets, education, and experiences. Two academic sector evaluators captured the essence of this narrative during interviews:

> ... so, this Values-engaged educative (VEE) approach kind of encourages evaluators to prioritize the voice of those often least or unheard. I think that is difficult to do in practice like it is in theory. So, I say in practice, I found that I always kind of talked to the key stakeholders, which are those who are like implementing the program. However, I have attempted to hear from the beneficiaries, intended beneficiaries, like students or their families. But I think that is done to a lesser extent, I will say, maybe compared to some other approaches, it is probably more, you know, it is important to make sure that the students' kind of know the results of the evaluation and kind of have a saying, I am giving feedback on various aspects, which is important. – Sade, Academic Sector Evaluator

Another academic professional discussed the creativity and value-laden interest she brings to board evaluation programs and projects. She emphasized that it is not enough for program organizers to just acknowledge the existence of marginalized groups, they need to evolve and continue to develop programs that are more inclusive of underserved communities. She summarized her thoughts as follows:

> So, I am not a content expert as it relates to different theories within science, mathematics, technology, or engineering. But what I am bringing is a perspective around evaluation, around working in educational programs. I am bringing experience and understanding around working with groups that are often—that have been historically marginalized in these contexts and trying to ensure that (a) their experiences are what they hope to get when they are part of a STEM program, and (b) that their voices are heard and represented so that program planners, going forward, know how to continue developing programs, given the big push to bring in folks who have been historically marginalized, and who have historically been left out or pushed out in STEM disciplines.
>
> – Tshepiso, Academic Sector Evaluator

A private sector evaluator provided a different perspective on the EDI Discourse by narrating her experiences through her identity and her role. Her story is typical, as the only person of color, and a member of an evaluation team comprising mostly white women working on an offshore, NSF-funded, community-based project involving URM groups of mainly black or Latino people. For the evaluator, navigating the tricky terrain of getting community buy-in and participation in the evaluation process is challenging enough, adding the extra layers of evaluation team members who are not from the community and who do not share the same or similar racial/ethnic backgrounds with them raises the barrier to the nth degree of complexity. In

situations like this, the NSF Broadening Participation agenda of EDI is brought into a sharp focus, and issues of trust and community relationship building become critical to the evaluation process. Under her role and identity, this evaluator has a position of power, which worked in favor of the evaluation overall; however, it was also fraught with major challenges, both internal to the evaluation as a team, and to the community they serve. Tlamelo, the evaluator, described a typical tension occurring within her evaluation team as the only person of color in a board meeting where the Broadening Participation agenda was on the table:

> ... I have been the only person of color in a room or in a meeting with a bunch of folk who are White and have them trying to tell me Like I'm bringing up a point as a person of color like this is for this project is when people of color and I'm a person of color, I'm saying, I'm like well you know that our folks are not like this like I'm your, I have a better, I think I probably have a better understanding than you do, but this is just is a very interesting dynamic to happen to navigate. – Tlamelo, Private Sector Evaluator

From the perspective of this private sector evaluator, discussions about Broadening Participation programs that are focused on equity invariably cause tension, more especially where representation in meetings of the community being served is low. According to Tlamelo, at times like this:

> It is always good to have an ally for yourself to have those conversations because there have been times when I've had to have one of my White female friends go, can all the White women, please hold, everybody else, including myself, included. –Tlamelo, Private Sector Evaluator

Taking a pause and reflecting on the moment not only helps diffuse the tension, but also teaches an important lesson "*why allyship is extremely important in this kind of work,*" Tlamelo opined.

When asked to provide further clarity on the concept of having an ally, and why it is critical to the work and practice of a STEM education evaluator, first, Tlamelo made the distinction about tagging equity and cultural responsiveness as the sole responsibility of Black people and deconstructed her idea of allyship from advocacy:

> I think that the thing about equity and culturally responsive work, especially, is that it can quickly become an onus on the people, on folk who are typically… it's ought to be part of this community, right? the Black folk? Okay, so Black people do a culturally responsive evaluation, so people will assume or think, or they're expected to do it.
>
> Tlamelo, Private Sector Evaluator

Second, Tlamelo explained the limiting position of lack of power and how it disadvantages people of color. In addition, she suggested that EDI issues should be the responsibility of everyone involved in the STEM education ecosystem to achieve the broader goals of NSF Broadening Participation. In providing a definition for the idea of an ally in STEM education evaluation work and drawing the different line to advocacy, which in evaluation is about relationships, she made clear that "allyship is about a true sharing of power." She described that as an advocate, an evaluator can speak up for an individual, participants, or groups of people in a community, and be "their voice" when they do not have one and when they are not available to speak for themselves. However, as an ally, she stated:

> I'm able to understand that let's not talk about it for you, but I'm going to get you in the room before you can speak for yourself, and I may speak for you when you're not there,

> but the first thing I'm gonna do is trying to get you in a room to speak for yourself, and then I'm going to be on and listen to what you have to say, and if is especially—if it's about your community like I'm not going to present my thoughts or feelings or expectations on you, so this is what you say you're needing; that's what you need. And I support you in that. – Tlamelo, Private Sector Evaluator

In essence, the definition of allyship as stated by Tlamelo is as follows:

> Allyship is about support, is about relationships, is about sharing of power more; so then I believe advocacy is because you can advocate for someone that's not there. – Tlamelo, Private Sector Evaluator

C. Attending to a STEM education evaluation by promoting the evaluator's values, expanding program effectiveness, improvement, and impacts

Promoting Evaluator's Values

By engaging program participants, diverse populations, or groups (e.g., students, women, parents, people living with disability, etc.); program managers; administrators; leadership teams; funders in projects; and evaluators in academic, private, and public sectors use their roles to help different stakeholders see the value of evaluation approaches and methodologies they embrace while also negotiating and balancing and being responsive to program needs. Findings from analysis of the interview data revealed that evaluators from all three sectors engage with the programming community by prioritizing their values. These values reflect in the form of approaches and methodologies used; for example, Eziamaka, an academic sector evaluator expressed how embracing the values-engaged educative and culturally responsive evaluation and the mixed-methods approach was useful in her line of work:

> One thing about culturally responsive approaches is that it brings in a lot of qualitative components. That allows us to get to the depth, right [?]... to get to those in-depth to look to understand the why behind some of the things we see and understand. But in my work, I also realized that working with STEM faculty, and STEM administrators, they are more used to quantitative stuff. And so, there's always—not always but very often—there is a tension where I've had some PIs say, "I don't know what to do with this qualitative stuff, right?" And so, my role then, as a mixed-methods person, or as a methodologist in that area, is to make the meaning right, and bring them along, right? To me, that's part of the values-engaged piece also… – Eziamaka, Academic Sector Evaluator

Along the same line, two other evaluators in the academic sector discussed using the values of their educational learning, training, and experience in surfacing and addressing issues in STEM education evaluation. One of them, Eghedo, captured the essence of *"the things that help"* in the program; he evaluated by telling the story of how he developed a theory of change and used a logic model to help navigate the evaluation process with stakeholders. Eghedo explained further:

> So, when you put the very end goal, say, what is it that you, really, really want to achieve? And they say this, then we work backward; we work backward, and as we are talking and going back and forth, you develop that theory of change and other linkages. And when you show them the theory of change to say, okay, this is what I had to say, this is what I had to say, is this correct? Now, why should X lead to Y? And they talk through that. And I put that all, you know, put all of that on the theory of change. So, there are no, well, okay, so as we are looking at the theory of change or your logic model, then we see

which components are going to be beneficial to different types of stakeholders… – Eghedo, Academic Sector Evaluator

Findings from the interview discussion with evaluators from the private sector are like the ones in academics; however, they also differ in some ways. One of the standout differences is the evaluation conducted by private sector evaluators in informal settings (e.g., museums, science centers, parks, gardens, etc.). In the light of bringing value to their work, several private sector evaluators describe the challenges encountered in the process of working in informal settings while having the goal of program effectiveness and a vision for long-term impact in mind. Findings from the analysis of the interview data revealed the challenges and lessons learned by Adaora, a private sector evaluator, encountered in using a previously validated survey in a different setting to engage the diverse group of citizens that patronize a national science festival in capturing their various experiences when attending the concert. Adaora describes the challenge and lesson learned:

> ...So, one interesting snag that we ran into with the engagement survey, is that it? It doesn't. We have done some other validation work where we thought, "Okay, we are taking this survey into festivals for the first time. It hasn't been validated there. We want to use it with kids and adults. It's never been validated with adults; like let's gather some data to see if the reliability and the validity seem to hold up in this different context and with these different audiences." And so, we had a history of doing that kind of work. And that ended up being lucky because in the context of this project, some of the questions you could just read and know like if somebody just watched a film, and they're talking about how engaging a film was, this question makes no sense. And so, we thought about

like, what are the teeny messages we could do to the language so that the integrity of the question we hope holds up? – Adaora, Private Sector Evaluator

Ofentse, the evaluator in the public sector added to the Discourse from a more divergent angle by focusing on issues of interest at the national level. One of the evaluation efforts discussed during the interview with the Ofentse entailed collecting annual or year over year kind of data to feed into a centralized database. Findings revealed that Ofentse values developing tools and technology that help to integrate data collected across programs and project evaluation sites nationally and feed them into a centrally accessible system for use by organizations. These data could then be used later on as a form of digital interventions to significantly improve STEM program outcomes across multisite evaluations where, for example, underserved populations or underrepresented communities (women, hard of hearing, native Americans, etc.) are located.

Expanding Program Effectiveness, Program Improvement, and Program Impacts

STEM education evaluators support systematic engagement in programs and projects intending to broaden the effectiveness, improvement, and impacts of current initiatives. Additionally, they seek opportunities to better understand prevailing evaluation contexts while coalescing diverse participants and stakeholders within a program community. They tend to prioritize discussions around barriers affecting underserved or underrepresented groups in STEM programs and utilize their evaluation approaches and methodologies to remove the perceived blocks or close access disparities. Findings from the interview data analyzed suggest that evaluators across the board in all sectors believe that program effectiveness is to be prioritized in the evaluation of STEM education programs.

One way this thread showed up in the analyzed interview data is when academic and private sector evaluators highlighted the thought process of disentangling evaluation data

collected along demographic lines to better understand participants' experiences. This systematic strategy that carries the promise of producing longer-term program effectiveness and improvement was described by Tshepiso, an academic sector evaluator:

> ...And so, when we're doing that, what we also do is we tend to disaggregate the data to look at okay, what are the experiences of female graduate students who identify as Hispanic, right? So, can we speak to their experiences? What are the experiences of Black male graduate students, as they are participating in this program? And so, that is one way in which we're trying to help. The program implementers or program managers understand what's happening for students collectively, but also within subgroups to know if things are not going well. Then, let's begin talking about how we do what's not working well, and how we begin to make changes, because what we don't want is to lose these students, and they say, "Because of this experience, I no longer want to be a part of STEM," right? So that's kind of one approach – Tshepiso, Academic Sector Evaluator

Regarding the expansion of STEM programs for long-term impacts, Rolake, who identified as working both in the academic and private sectors, discussed how painstaking and challenging it is to build and develop programs before the idea of impact can begin to take on tangible meaning. Rolake then rejected the notion of scientific methods to show measurable outcomes of program evaluation. Rolake opined that "*It was a bit unfair, in my opinion, you know, to say that you have to show impact and what works clearinghouse when it takes two to three years you know just to get the program up and running efficiently***."** Rolake then drilled down her point on mixed methods while rejecting emphasizing scientific strategy in this way:

> **...**so that is not the type of work that I eagerly seek, let's put it that way. I've done it, but that is not the kind of experience… at least I'm going to keep it to that experience. And

> felt like the evaluators were less impactful in terms of helping the client with the process and formative evaluation because we had to be so focused on metrics; I mean we even had to go to a separate conference just for that. – Rolake, Private/Academic Sector Evaluator

Theme 2. STEM education evaluators had positive things to say about professional development initiatives and community organizing practices

Professional Development

STEM education evaluators play a crucial role in evaluating the impact and effectiveness of STEM education programs and initiatives. To achieve their goals, STEM evaluators described efforts to support their professional growth and success during the interviews. They mentioned several professional development opportunities, networking opportunities, and community-building and organizing experiences in their line of work. Their educational backgrounds range from degrees in psychology and sociology to higher education leadership and policy studies, design and systems thinking, and teacher education training. Interview participants derive benefits from attending workshops and conferences. Seven of the interview participants have AEA STEM TIG affiliation, an additional five have attended AEA conferences in the past, and the majority (92%) suggest they have attended courses that are specifically designed for evaluation professionals. These opportunities allow evaluators to stay current on evaluation approaches, methodologies, and best practices, and to learn from experts in the field. As Tshepiso puts it:

> I am very much involved with the [AEA] STEM topical interest group. And we meet regularly, which is a great place to also network with other STEM evaluators; it's a great place to learn about the work that they're doing to provide resources for one another, and

> also to just brainstorm where we see the field going in the coming years. – Tshepiso, Academic Sector Evaluator

Other evaluators echoed Tshepiso's opinion that networking is a key component of professional development for STEM education evaluators. In addition to attending conferences, workshops, and online events to connect with peers and build relationships with others in the evaluation community, STEM evaluators work collaboratively with others on projects, share resources and information, and develop a support network. Another evaluator, Eziamaka, shared:

> One thing I would say is that I have become more of a coach, and I don't know the right word for it. But as teams have struggled through COVID, I have found that I have taken on the role of supporter and cheerleader just to like to encourage or maybe chief encourager—a role that I've taken with all the teams… – Eziamaka, Academic Sector Evaluator

Community Building and Organizing Practices

Building a community of STEM education evaluators is another important thing that can help to increase the visibility and impact of evaluation in STEM education. Interviewees reveal in dialogue that part of their work involves participating in local, regional, or national organizations focused on funding programs and implementing evaluation of STEM education initiatives. These organizations provide opportunities for professional development, networking, and collaboration, as well as advocacy and representation for the evaluation profession.

As with many others, most of the interviewees stumbled into the field of evaluation through their work in academic institutions and private consulting businesses or public/governmental agencies. They do a lot of STEM evaluation with organizations like the

Department/Board of Education, the Department of Defense, and K–12 rural and urban schools, and are funded by the NSF, NIH, or private granting organizations. They extend their research and practice of evaluation approaches deeper than the cultural and contextual factors and aim to reach people at the grassroots. Mpho, an evaluator in the academic space, explained it further:

> And so instead of thinking about just being culturally responsive, I want to be a community engaged. And so that looks a little bit different, depending on which project I'm doing. So, I can kind of explain what that looks like in my world. But instead of doing the evaluation, where I kind of come in, in the end, I might help folks design research, or work with my team to design research with the community, not for the community, with the community. And that means that their research looks a little bit different. And then the evaluation of that research looks a little bit different than perhaps a traditional research study. – Mpho, Academic Sector Evaluator

Another evaluator, Tshepiso, added voice to the context of community building by allowing prevailing climatic factors to inform and drive their evaluation approach and methods for a given project. In her words:

> And so, for instance, with the elementary school that I was just referring to, we are allowing the families and the students to be the drivers of how the program is implemented because then that frames or informs the questions that are going to be asked in the evaluation; that frames or informs the data that will be collected, right? That's going to frame and inform us even going and disseminating the finding back to these parents and these families and, also, the surrounding community. So that is very much a big driver of this work. – Tshepiso, Academic Sector Evaluator

Another dimension of community building and organizing practices that showed up in the interview analysis is mentoring, where participants actively seek and engage in professional relationship building with a more experienced practitioner/mentor or less experienced colleague or mentee. This scenario often translates to evaluators' practice, where they get a chance to evaluate the implementation of mentorship programs and initiatives. Ugochukwu, one of the private sector evaluators, describes her involvement in a project as follows:

> For other programs, whether the Department of Ed. funded, or NSF funded, that is much broader with their possible uses of funding, and broader evaluation budgets; it's much easier to have the time to explore student-level outcomes to see if the intervention at the classroom level is impacting all students or if one demographic is left behind. We have found in a few instances that Hispanic students, for example, might have a lesser benefit from a peer mentoring opportunity, if you will. This is one example. If it wasn't a good fit with the peer mentor. – Ugochukwu, Private Sector Evaluator

Additional items referenced in the discussions include competency and skill building in STEM education evaluations. Evaluators noted the process of acquiring the knowledge, skills, and attitudes necessary for effective practice in STEM settings. Academic sector evaluators were more detailed with content knowledge in their specific STEM fields. Some cited pedagogical knowledge for teaching STEM and technology skills for using digital tools in the classrooms. Private sector evaluators were more detailed about *Shepherding* in their STEM evaluation practice.

Adaora specifically describes how shepherding helps in her work in providing support and guidance to informal STEM learners as they navigate the STEM education landscape. This includes organizing and helping learners and program stakeholders within spaces such as

museums, science clubs, citizen science, parks, subways, etc., come together and build a viable STEM community. Adaora believes the future of STEM education evaluation approaches and methodologies in the context of informal settings is going to be decided by how much investment evaluators are willing to deposit into shepherding.

> And so, I hope that we are going somewhere that looks different because it involves much more shepherding and thinking with not just like, what should my evaluation method look like? But what should your relationship-building method look like with the Client? What should your partnership methods look like, and how do we think about whether you're doing the due diligence around authentic work like authentic community-based work? – Adaora, Private Sector Evaluator

Theme 3. The era of accountability has evolved, as seen through the lenses of leadership, social justice, and self-socialization

The following sub-themes have been identified in the analysis of interview data and used in the presentation of findings for the theme.

Accountability from a Leadership Perspective

STEM education evaluators made clear during the interview discussion that the era of accountability in STEM education is here, and programs/agencies are beginning to embrace it. Concomitantly, this new era is characterized by a growing focus on evaluation, with increasing attention to leadership, social justice, and self-socialization. In this context, STEM evaluation is becoming an increasingly important field, with new funders and clients driving the development of new, client-based, or need assessment/evaluation approaches and methodologies. As Sade, an academic STEM evaluator, sees it, the current state of "*the field has irreversibly changed*" due to the Covid-19 pandemic, and the future of STEM evaluation is open to opportunities made possible only

by the outbreak of the infectious disease caused by the SARS-CoV-2 virus. Sade expounds on the need to be accountable:

> And, this idea of new funders, you know, as we continue to be in the era of accountability, more and more people are funding various projects, especially within STEM. And so, there are some new players, philanthropic organizations, as well as, you know, that the key government organizations continue to be kind of committed to evaluations. – Sade, Academic Sector Evaluator

Accountability from Social Justice and Equity, Diversity, and Inclusion Lenses

Accountability is becoming an increasingly important factor in STEM evaluation, as leaders in the field seek to understand the impact of their programs and initiatives on the individuals and communities they serve. This is particularly relevant in the context of social justice, where STEM education evaluation has the potential to play a significant role in addressing equity concerns while embracing diversity and promoting the inclusion of a wider audience in the STEM ecosystem. The Discourse on accountability surfaced the interests and particular evaluation approaches that practitioners' care about. Findings show that interview participants expressed personal interests as a human in space around the outcomes of target populations and professional interests in projects or research interventions they pursued. Eziamaka, an academic sector evaluator, expands the conversation:

> And the team that hires us knows where we stand up front and where we're able to articulate that so that there are no issues along the way because then we're both agreeing that "Yeah, we care about this; we understand this is an approach, and we want to be held accountable to this." And so, there is that I think it opens up the door for that dialogue upfront, and then that accountability going forward. And I've so far in my work, I found

that it's a strength, and it helps the project move further along the path. – Eziamaka, Academic Sector Evaluator

Others from the academic and private sectors added to the accountability Discourse on STEM Education Evaluation. They all agreed that there is a shift toward more equity-focused proposals and grants supported by STEM education funding agencies and leadership. For example, Rolake, an academic/private sector evaluator suggested that she and her colleagues have experienced "...three to four NSF proposals with potential clients, and all of those proposals from NSF had an equity focus." Furthermore, another private sector evaluator, Latifah, observed a marked shift in NSF funding priority towards centering equity:

> And then I have noticed that even funders are beginning to think differently about funding projects that center equity, right? So, like even NSF, this year, started a new grant funding program that would focus on essentially like, you know, informal science learning and centering equity. – Latifah, Private Sector Evaluator

And then, in terms of decisions about evaluation methods and methodologies, Sade, an academic sector evaluator, added to the accountability Discourse:

> I always start with evaluation questions. And I always make sure that a question, there is a question about diversity, equity, and inclusion. And so, by including this question, I am trying to make it explicit. I also try to have kind of informal conversations and informal commitments to bring up issues of diversity and equity with stakeholders and other kinds of aspects of the programming. And so, I make decisions—those decisions are kind of based on my own personal, ethical, and moral values as a professional evaluator. – Sade, Academic Sector Evaluator.

Accountability from the Viewpoint of Self-Socialization

The sheer sense of injustice that is reverberating throughout the nation in the wake of George Floyd's murder on May 25, 2020, prompts a resurgent debate on racial reckoning across disciplines without leaving STEM education out of it. The Discourse on accountability from the standpoint of self-socialization emerged from the interviews with STEM evaluators, with the majority referencing George Floyd's killing as a watershed moment for personal and professional reflections. Findings in the interview data further reveal discussions on accountability in the context of racial reckonings and antiracism debate. Considering their evaluation practice, interview participants expressed accountability as a process of taking responsibility for their thoughts, beliefs, and actions, as well as the impact they have on others, especially regarding race. This process involves self-socialization and self-education, as the STEM evaluators seek to expand their understanding and awareness of the role of race in society, in practice, and in their personal lives. Sade, an academic sector evaluator captures the essence of the self-socialization and a new sense of duty in the aftermath of George Floyd's death and its potential implications for the current and future of STEM education evaluation:

> I think many within the field are excited about and applaud these new commitments to diversity and equity within our fields. And I do think that I do see that, now that we are kind of almost two years out. I do see that as continuing, maybe not as strongly, but I still see it as a priority for many people…So, I think there's kind of this commitment to, at least, educate ourselves about issues of anti-racism and Equity, Diversity Inclusion. I think people are trying to engage with these topics in a way that is more than just utilizing them as a buzzword, in a way that it is more than just thinking about it from a

kind of shallow perspective. They want to meaningfully engage with these topics. And I think that will continue. – Sade, Academic Sector Evaluator

Further, findings showed that race, class, and gender intersect and shape the experiences of STEM education evaluators, and critical race theory showed up in conversations. This highlights how racism is ingrained in the fabric of society at large, perpetuated by laws, institutions, and cultural norms, and permeates the field of STEM education evaluation. Interview participants describe steps taken to build awareness, which is crucial in the pursuit of accountability and antiracism, as it helps individual evaluators understand how they may unconsciously perpetuate or benefit from systemic or historical racism in their line of work. Some of the evaluators described their thought process in grappling with the issue of accountability from the viewpoint of self-socialization in their praxis:

> So, I have been working to continuously evolve my thinking around these topics. Because I used to define groups as historically underrepresented or historically marginalized. It is history, but it is not just that it is historically, but systematically, right. Like, for people, this is on purpose that certain groups of people do not have access to resources, certain groups of people are unable to have certain experiences, and certain groups of people do not feel included. – Sade, Academic Sector Evaluator

Sade further shared her effort to expand reading and understanding of the underlying topics of racism "*from a historical perspective, outside of the field of evaluation."* Reading materials beyond the evaluation field is helping to "develop a deeper understanding of the historical context" and "*understand that this idea, that it is not, as a system, broken; the system is working exactly as it was designed.*" Mpho, another academic sector evaluator, took the

discussion a step further and drilled down with an example of a project she handled that required her to be more responsive in practice:

> We need to be responsive in that we hear and see what people are trying to communicate to us. And we pivot. So, we completely changed our design. And we started stratifying the focus groups, by gender, and by race, and so that we could try to get to a place where people would have a comfortable, safe—where they didn't feel like they had to go out on just the chat so that they could effectively communicate in a larger group. Because we wanted to hear what they had to say. And we wanted to create a comfortable environment. So, I think being able to pivot is important. – Mpho, Academic Sector Evaluator

Interviewees also shared that the social and political environment plays a crucial role in shaping the conversation around accountability and antiracism. Policy changes, media representation, and activism can create opportunities for education, growth, and systemic change. However, it is ultimately up to individual evaluators to take personal responsibility for their actions and strive for accountability in their evaluation practices and personal spaces.

Theme 4. Data Utilization in the Age of New Technology is significantly influencing how STEM education programs are evaluated today and for the future

The role of technology and big data is significantly influencing how STEM education programs are evaluated today and in the future. The interview discussions with STEM evaluators surfaced many topical issues, chief among them the use of evaluation data concerning new technology. Other items mentioned by STEM evaluators in conversations are Hybrid evaluation, virtual forms of data collection, and in-person or face-to-face STEM evaluation work.

Data utilization in the age of new technology is having a significant impact on STEM education programs and the ways practitioners view evaluation approaches and methodologies. With advancements in technology, STEM evaluators are presented with new options for access to data, methods of collection, processing, and analyzing large amounts of data sets. This era of new technology has ushered in multiple ways that are helping evaluators with tracking program implementation and challenges, as well as stakeholder participation and experiences.

Furthermore, through data utilization, STEM evaluators can monitor program effectiveness, improve educational outcomes for students and program participants, report on progress and successes to program managers, and inform funding agencies for decision-making. Consequently, the current landscape, and even more so, the future direction of STEM education evaluation, is anchored in new ways of imagining and engaging with data. Sade, an academic evaluator, put it in perspective:

> This idea of thinking about data, and big data, is a huge thing. And utilizing data, we create so much data on our phones and our everyday lives and our ZOOM meeting, so the evaluators [are] figuring out how they can responsibly and ethically tap into that data; I think is the direction the field is gonna continue going. – Sade, Academic Sector Evaluator

In other areas, private and public sector evaluators are using data analytics tools to collect data to address complex challenges facing teachers, administrators, and other stakeholders in Higher Education, Departments of Education, and Federal Agencies. In these settings, they use approaches that work by bringing together multi-sectoral stakeholders to build shared knowledge of the systems and then find opportunities to leverage change across the systems.

In one of the interviews, Adekunle, a private sector evaluator, described the process of *Ripple Effects Mapping* a participatory research tool, to actively work with K–12 educators in a summer camp to collect the untold stories of the teachers and behind-the-scenes activities that could ripple out from the summer workshop. Adekunle explains:

> I have several STEM projects that have been involved with—that have engaged science teachers from mostly, I think, middle school and secondary science teachers in kind of professional development projects over the summer. And just as an example, in one case, that is, one group of teachers [worked] together on STEM-related learning throughout the three summers, and so there was a lot of, you know, they, they would learn something in the summer and then do something with the following year, see how it worked with their students to reflect on it, and then come back the next summer and, and talk about what they've done and what they learned. And at the end of that three-year cycle, we did a ripple effects mapping process. And, you know, the point of ripple effects mapping is really to talk about is kind of rooted in, in appreciative inquiry. – Adekunle, Private Sector Evaluator

Technology plays a crucial role in evaluating STEM education programs and projects, and for evaluators in the public sector, it is not just a tool but is an integral part of the evaluation practice as it enables conversation starters, facilitates, and smoothing issues during meetings and is key to data integration in central databases. Ofentse, the public sector evaluator narrates his viewpoint on technology and how it is helping to shape his relationships with clients and co-workers across the agency in which he works and current evaluation practice in the STEM field:

> We have to talk every couple of months, you know; we have to constantly engage. So, it's a way of still building a relationship, you know; it's another tool to building a

> relationship with a client, essentially. And so, every meeting we have about the system is a meeting about evaluation. You know that's, it's, that's kind of how I think about it. So. And vice versa. Sometimes we're having a more general evaluative conversation, and we talk about technical things like, "Oh, actually, we should talk about how we want to collect that data system," or "How do we tag that data?" So, there are some programs I've been working with because it always goes back and forth in that regard. And so that's how, you know, for me, technology is not just a tool, but it is part of my practice, in a very kind of deep sense. Like, that is how I operate as an evaluator. – Ofentse, Public Sector Evaluator

There were additional conversations around technology that enabled the use of hybrid evaluation methods. This combines traditional in-person evaluations with virtual forms of data collection. Evaluators suggest this approach provides a more complete picture of program participants' and stakeholders' experience in a way that allows for a wider range of data sources to be used in evaluations. Ugochukwu, a private sector evaluator, commented:

> I do think we're going to see more hybrid evaluations with a very, very intentional focus on the use of online and face-to-face observations, a lot more virtual and asynchronous; even data collection, the sharing of student data files, I think that's going to be much more prominent. – Ugochukwu, Private Sector Evaluator

As mentioned in the dialogue with the STEM evaluators, below are other variations of data utilization in the era of technology influencing the Discourse of STEM Education Evaluation that has implications for the future.

- **Virtual Forms of Data Collection:** The use of technology has also enabled virtual forms of data collection, such as online assessments and quizzes, which can be used to

supplement in-person evaluations. This allows for greater flexibility in the evaluation process and enables data to be collected from a wider range of program participants, regardless of location.

- **In-Person or Face-to-Face STEM Evaluation Work**: Technology has also made it possible to conduct STEM evaluations in-person or face-to-face, using virtual tools such as teleconferencing and online collaboration tools. This enables educators and evaluators to work together in real-time, regardless of location and facilitates the sharing of information and data in real-time.

Essentially, technology is playing a crucial role in shaping how STEM education programs are evaluated today and in the future. These advancements in technology are enabling more comprehensive and valid evaluations, as well as providing greater flexibility and convenience in the evaluation process.

Theme 5. Discussions on approaches and methodologies are leading to the next frontier in the evaluation of STEM education programs

There is a growing list of evaluation methodologies, approaches, theories, models, frameworks, and strategies available to STEM evaluators. As revealed in this study's findings, as STEM evaluation grows and expands, so grows the list of available options of evaluation approaches and methodologies in the field due to old and new challenges faced by programs and projects, varying degrees of complexity of problems encountered, values and interests of evaluators, as well as needs and demand of funders and program organizers.

Collaborative Evaluation Approaches and Mixed-Methodologies Dominate the Discourse

Evaluators discussed a wide range of evaluation approaches and methodologies and suggested that evaluation questions drive the choice of the approach taken and inform the

rationale behind the methods deployed. Ugochukwu, a private sector evaluator, summarizes a positive outlook on the STEM evaluation field's current and future state:

> I'm optimistic that given the requirements of contextualizing—everything went wrong during COVID—that people will realize it's not just about the numbers, it's about the story. What was the narrative behind what was going on? What worked, what didn't work, you know, what we tried. So, I'm optimistic that we'll come out of this with more appreciation for qualitative data and mixed methods approaches. – Ugochukwu, Private Sector Evaluator

The STEM evaluation field is shifting toward centering equity in all areas of evaluation endeavors

The field of STEM evaluation is undergoing a major shift toward equity-centered evaluations. This means that evaluations are being conducted to ensure equal access to STEM education for all groups of people, including those who have traditionally been marginalized, underserved, or underrepresented. STEM evaluators highlighted that, in addition to collecting data, equity-centered evaluations also involve engaging with stakeholders from diverse backgrounds to ensure that evaluations are culturally responsive and reflective of the experiences and perspectives of all groups. This helps to ensure that evaluations are truly inclusive and that recommendations for improvement are grounded in the experiences of those who are most impacted.

Evaluation Question 2a: To what extent do the initial findings from the STEM evaluation literature of Discourse resonate with STEM education evaluators about the current and future state of the field?

Data Used

The data used to answer this question included: (a) STEM evaluator interviews, (b) document review items—AEA STEM TIG evaluation abstracts, and (c) STEM education evaluation literature.

Overarching Finding

The question in this section was answered by examining the STEM education evaluation Discourses regarding the approaches and methodologies utilized by evaluators within the STEM ecosystem. The analysis focused on how STEM evaluators characterize Discourses connected to evaluation approaches (e.g., values-engaged, educative; culturally responsive; utilization-focused, responsive, and so on) and numerous methodologies which justify the choice of methods used (e.g., case studies, surveys, focus groups, interviews, etc.).

The extent to which initial findings from the STEM evaluation literature of Discourse resonate with STEM education evaluators about the current and future state of the field differs by the sector in which the evaluator is operating or practicing. To provide clarity, the findings have been organized by evaluation approaches and methodologies and compared by disaggregating the interview data by the three sectors (a) academic, (b) private, and (c) public. In addition to the interview data, the AEA STEM TIG abstracts were disaggregated along the three sectors, and both were compared to the initial findings from the evaluation literature.

- **Discourses of Evaluation Approaches in the Literature compared with the findings from STEM Evaluation Interviews**

The initial findings for the evaluation approach in the literature were drawn from the work of the following authors (a) Stufflebeam and Coryn (2014); ((b) Greene, et al. (2011); (c) Hood, et al. (2015); (d) Lawrenz and Huffman (2004); (e) Coryn (2004); (r) Newcomer, et al., (2015); and (g) Huffman et al. (2006).

The evaluation approaches listed below was chosen from the work of Stufflebeam and Coryn (2014) because they have been described as "the best and most applicable of the program evaluation approaches" in the evaluation literature. They include (a) client-centered/responsive, (b) utilization-focused, (c) decision-and-accountability-focused, (d) consumer-oriented, (e) constructivist, (f) case study, (g) outcome/value-added assessment and accreditation, and (h) a deliberative democratic approach. The Values-Engaged Educative approach was picked from the work of Greene, et al., (2011) because of their explicit description of its uniqueness and fitness for STEM education program evaluation. Hood, et al., (2015) Culturally Responsive Evaluation were chosen for their descriptions of culturally responsive evaluation's rising popularity and currency in the evaluation field.

The Collaborative/Partnership approach was adapted from the works of Lawrenz and Huffman (2004) and Cousins and Chouinard (2012) for depth of definition and clarity of description. The section on building evaluation capacity was added from the works of Stockdill, Baizerman, and Compton (2002), and Huffman et al., (2006) because the issues addressed therein resonate with the core purpose of this study.

- **Evaluators in the STEM fields are most interested in Client-Centered/ Responsive and Utilization-Focused Evaluation Approaches**

Findings in the interview data seemed to suggest that evaluators from all three academic, private, and public sectors discussed the idea of engaging in client-centered, responsive, and utilization-focused evaluation approaches in STEM education programs (Table 17 and Figure 6). A common thread that runs through the interview data is the fact that all the STEM evaluators interviewed described working with and providing support for diverse participants, groups, or clients, including students, teachers, faculty, program, and project managers, URMs (women, people with disabilities, etc.), principal investigators, citizens in science and within the community, and program funders and agencies. Following are sample quotes during interviews with the STEM evaluators that show ways they engage, interact with, and respond to the evaluative needs of their various clients and stakeholders in the programming community.

> This is a program that's in a higher ed context, right? It has multiple components around high school student experiences, undergraduate student experiences, graduate student experiences, K through 12, and teachers. And then like, the faculty and staff of this, who like to interact and manage the program. And so given, like the different subgroups that I just identified, and knowing that this project has a focus on ensuring that experiences are inclusive, ensuring that there's a diversity of individuals who can participate, and that experiences are equitable, what we've been doing with them regularly is not just, okay, understanding how well things are working from, like, the effectiveness or start over what aspect of understanding the effectiveness of the program, what we've been doing, or climate studies with this program, to dig deeper into what is happening. – Tshepiso, Academic Sector Evaluator

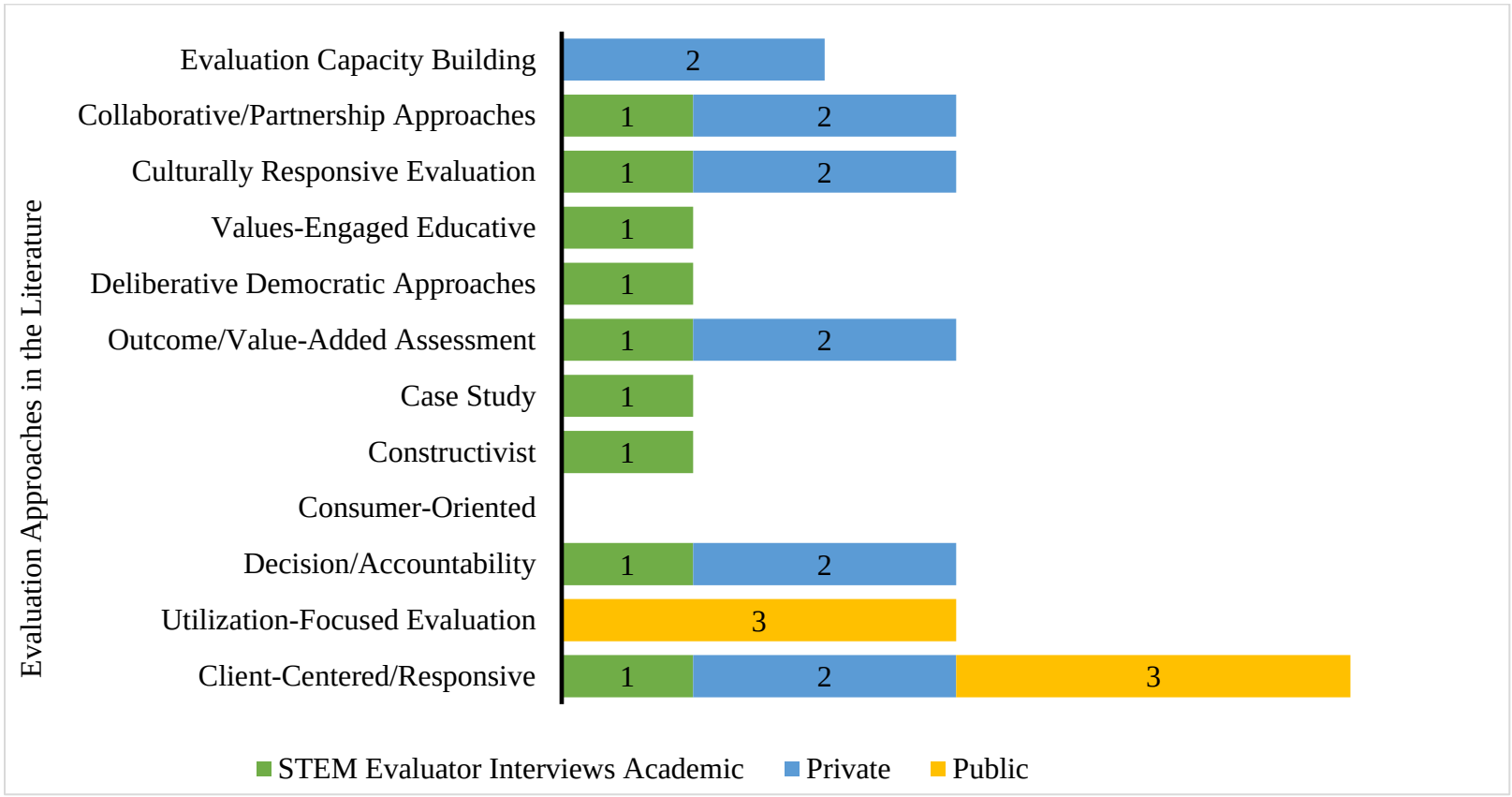
So, like anything I've ever learned, from the literature about evaluation theory, evaluation methods, evaluation, and research, all of that happened on kind of a need-to-know basis, because I consider myself, like, I consider myself a practitioner who publishes, not an academic. And so, it's very much that practitioner, like, "How is this useful to me? Is it useful to my clients?" If so, then let's think about it and if not, then I don't have time for that… Adaora, Private Sector Evaluator

Table 17. The Discourse of the Current and Future State of the STEM Evaluation Field Regarding Evaluation Approaches (STEM Evaluation Practitioner in Academic, n = 6; Private, n = 5; Academic/Private, n = 1; Public, n = 1)

Evaluation Literature	**AEA STEM TIG: Abstracts**			**STEM Evaluation Practitioner: Interviews**		
Approaches	Academic	Private	Public	Academic	Private	Public
Client-Centered/Responsive[1]	✓	✓		✓	✓	✓
Utilization-Focused[1]	✓	✓	✓			✓
Decision/Accountability[1]		✓	✓	✓	✓	
Consumer-Oriented[1]			✓			
Constructivist[1]				✓		
Case Study[1]	✓	✓	✓	✓		
Outcome/Value-Added Assessment[1]	✓	✓	✓	✓	✓	
Deliberative Democratic approach[1]				✓		
Values-Engaged Educative[2]	✓		✓	✓		
Culturally Responsive Evaluation[3]	✓	✓	✓	✓	✓	
Collaborative/Partnership	✓	✓	✓	✓	✓	
Evaluation Capacity Building	✓	✓	✓		✓	

[1]Stufflebeam, D. (2001); [2]Green, J., Boyce, A., & Ahn, J. (2011); [3]Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015)

Figure 6. STEM Evaluation Approaches Discussed in the Literature Compared to Perceptions of STEM Evaluators in Academics (1), Private (2), and Public (3) Sectors



> You know, the support process of really getting people out of denial was my initial practice; there was, like, kind of working people through a denial phase of understanding that, you know, they were like, well, you know, this was March 2020, "Oh, we'll just, you know, we're gonna reopen, you know, in the summer." "Oh, we 'll reopen in the fall." And then, you know, all were reopened next year. And so, I just like, yeah, work, you know, getting people to accept the reality of today was a lot of lot, I had to shift into that. And then, and then support them and say, "Look," you know, and then, you know, and then they can, you know, say, "It's not working," is like, well, there's, you know, I can't say like, "Look, there's, you know, you're—It's you; you're not seeing it right now. And so, until you see it, it's not gonna, you're not going to see some benefits here." So. So yeah, kind of, I would say that that's, that's kind of the primary one. – Eghedo, Public Sector Evaluator

Other findings that emerged from the data analysis regarding evaluation approaches include the following:

Evaluation Approaches from STEM Evaluation Literature versus STEM Evaluator Interviews

1. Academic and private sector evaluators have more in common that resonates with initial findings from the literature. The four evaluation approaches from the literature that resonates with both academic and private sector evaluators include (a) Collaborative or Partnership Approaches, (b) Culturally Responsive Evaluation, (c) Outcome/Value-Added Assessment, and (d) Decision/Accountability (Table 17 and Figure 6 above).

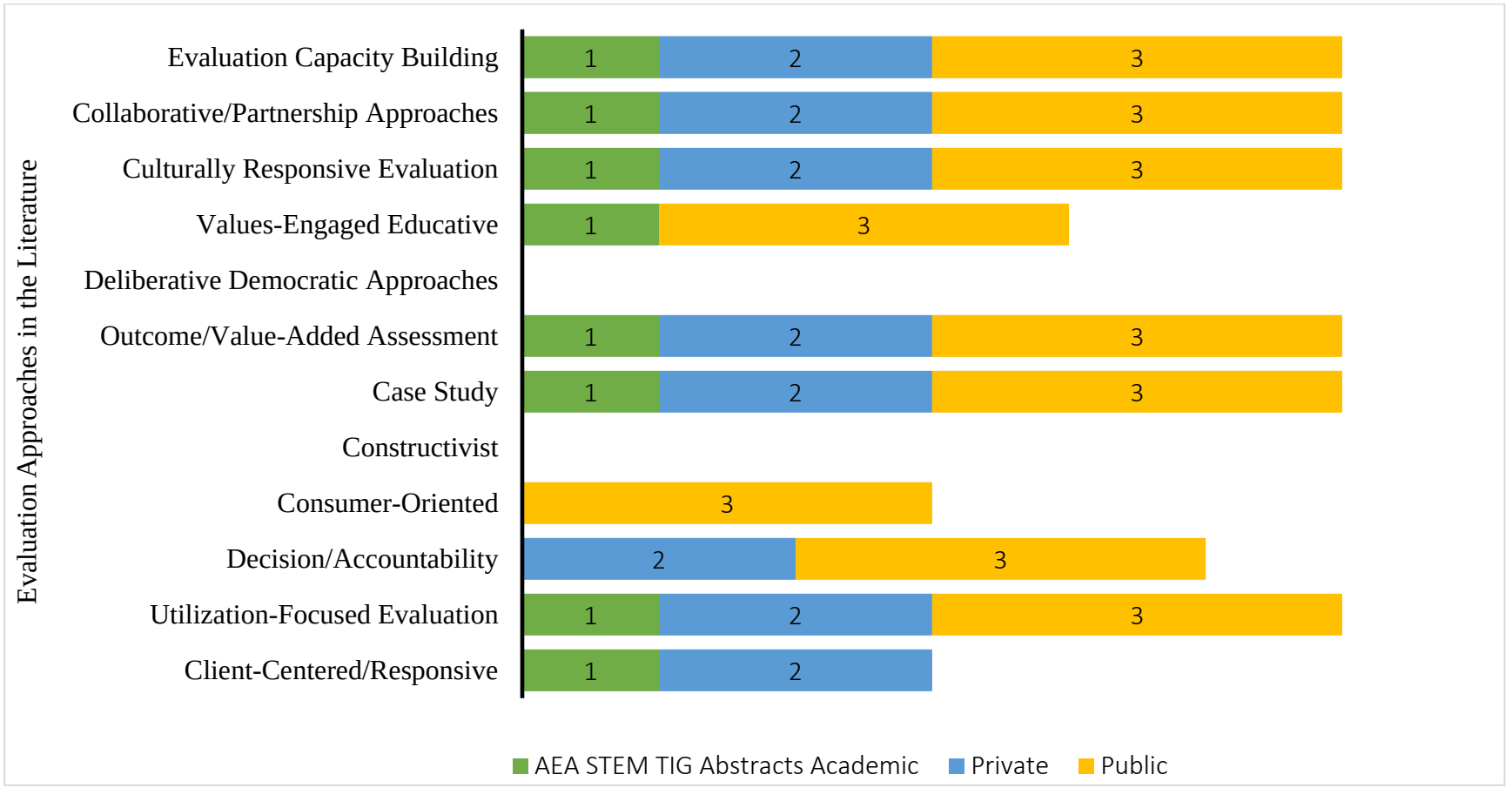
2. Four of the evaluation approaches from the literature resonate well with only the academic sector evaluators. These include (a) Outcome/Value-Added Assessment, (b) Deliberative Democratic Approaches, (c) Constructivist, and (d) Case Study.
3. The evaluation capacity building (ECB) from the evaluation literature resonates well with only the evaluator from the private sector.

Evaluation Approaches from the STEM Evaluation Literature versus AEA STEM TIG Abstracts

Different from the STEM evaluators interviewed, where only two approaches resonate with them in the evaluation literature, evaluators in the STEM TIG abstracts discussed at least six evaluation approaches that are like the evaluation literature. These include (a) Culturally Responsive Evaluation, (b) Collaborative/Partnership or Partnership approaches; (c) Evaluation Capacity Building, (d) Case Study, (e) Outcome/Value-Added assessment, and (f) Utilization-Focused evaluation (Figure 7 below). Other findings in the AEA STEM TIG abstracts include:

4. Values-Engaged Educative evaluation approaches from the literature resonate well with academic and public sector evaluators.
5. Decision/Accountability approaches from the literature resonate with private and public sector evaluators.
6. Client-Centered/Responsive approaches from the literature resonate with academic and private-sector evaluators.
7. Consumer-Oriented approaches from the literature resonate with only the public sector evaluators.
8. Constructivist and Deliberative Democratic Approaches from the literature did not resonate with any of the evaluators in the STEM TIG abstracts.

Figure 7. STEM Evaluation Approaches Discussed in the Literature Compared to Perceptions of STEM Evaluators in Academics (1), Private (2), and Public (3) Sectors



Discourses of Evaluation Methodologies in the Literature Compared with the Findings from STEM Evaluation Interviews

The initial findings for the evaluation methodologies in the literature were drawn from the work of Lawrenz and Huffman (2006). They provided a comprehensive and rigorous discussion for thinking through the process undergirding the varying and evolving choice of methods available to STEM evaluation practitioners. The evaluation methodologies outlined by Lawrenz, and Huffman (2006) used in the analysis of this study include the following:

1. **Quantitative Designs**
    a. Randomized Control Trials, RCTs
    b. Quasi-Experimental Designs
    c. Correlational Designs
2. **Qualitative Methods**
    a. Case study
    b. Status and Survey
    c. Site visits focus groups, individual interviews
3. **Mixed Methods**
    a. Mixing of philosophies
    b. Research designs
    c. Data collection devices

Evaluators in the STEM Fields Provide the Most Justifications for Data Collection Devices, and Status and Surveys for Evaluation Methodologies

Outlined below is a set of findings from the analysis of the interview data, with additional findings from the AEA STEM TIG abstracts, both used in comparison with the initial findings from the evaluation literature. The findings revealed that mixed methods (data collection devices), and qualitative methods (status and survey) were the two most frequently discussed evaluation methodologies during interviews with STEM evaluators. An example of a mixed method thought process is highlighted in the discussion with an academic evaluator during an interview, who opined those mixed methods are valuable for the future of STEM program evaluation given the social justice issues the nation is currently facing:

> So, I think we are going to be seeing more evaluators who take approaches that are justice and equity-oriented, right? And as a result, they'll be using, I would say, a diverse number of methods, along with those approaches. As someone who uses mixed methods and teaches mixed methods, I think there is a lot of value to integrating mixed methods in STEM education evaluation, where approaches are drawing from equity and justice. And so, I think we're going to be seeing more of that. And there are some scholars like Donna Mertens, who talk about the use of, you know, you can use both quantitative and qualitative methods and still have an equity orientation. And so, I think that's where we're going to see the field going, as we think about approaches and methods—which I think is a good thing. – Tshepiso, Academic Sector Evaluator

Other evaluators in the field attribute their choice of methods to thinking deeply and responsively toward program participants' needs and learning environment. As one other evaluator from the academic sector suggests:

So, projects that deal with students, right, so where it's STEM Education, we're dealing with students in classes, or students and workshops and those kinds of things.

We are always going to see that mix because you want to get the pre-post kind of thing. So, we'll see survey use there. And that resonates with folks, right? But I have one project where we have that, and so we have pre-post surveys, but response rates were extremely low, right? They tried to survey even the faculty mentors, and response rates were kind of confusing, not no [sic] response rates were poor, but the survey results were. They were kind of confused, and at times like that, we got to say, "You know what, let's try a different approach," or "Let's complement or supplement this with some interviews, right? Or some observations or document reviews." – Eziamaka, Academic Sector Evaluator

Other findings that emerged from the data analysis regarding evaluation methodologies include the following:

Evaluation Methodologies from the STEM Evaluation Literature versus STEM Evaluator Interviews

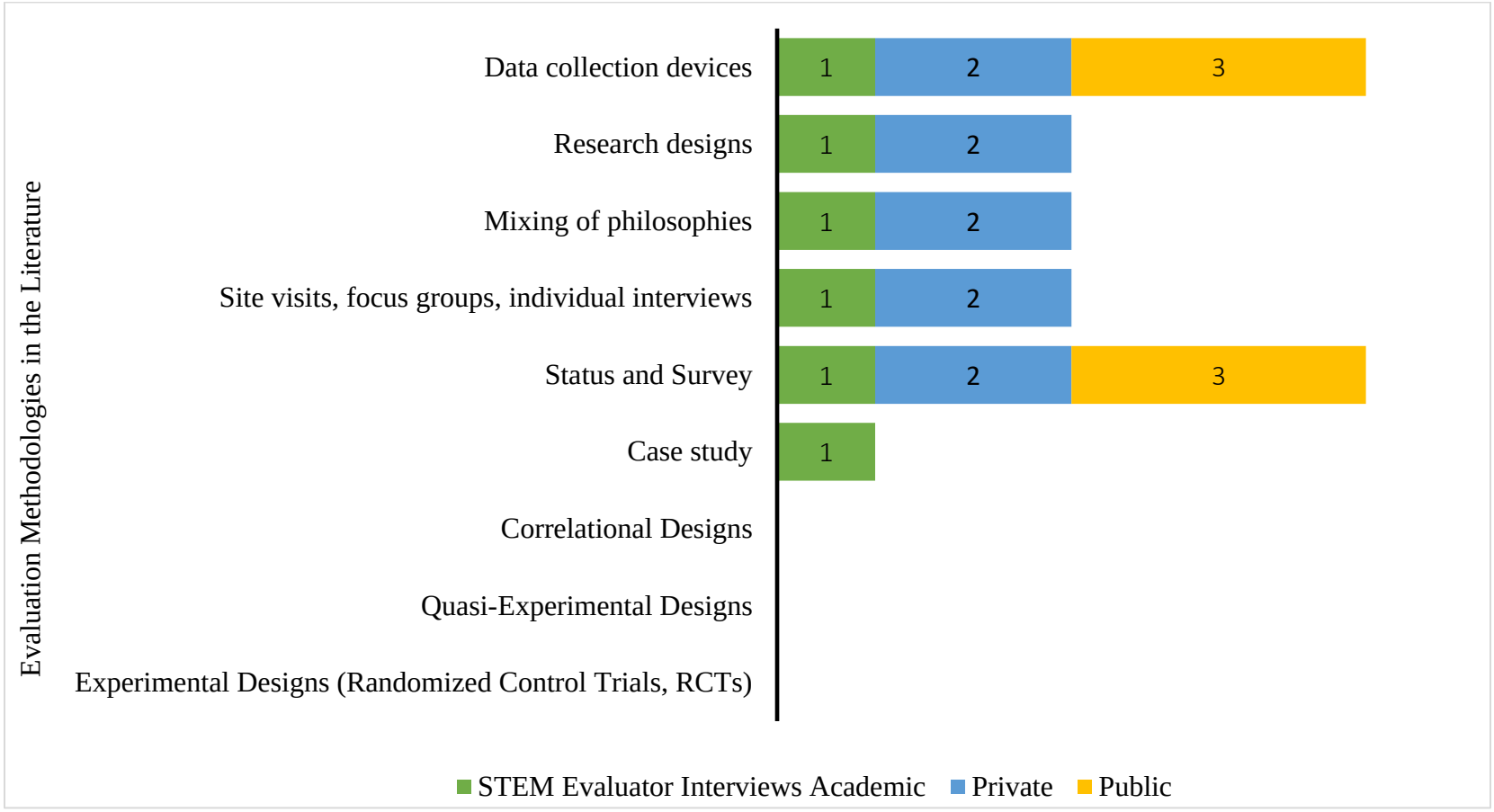
Similar to evaluation approaches, academic and private sector evaluators have more in common that resonate with initial findings from the literature regarding evaluation methodologies. Three evaluation methodologies from the literature that resonate with both academic and private sector evaluators include (a) research designs; (b) mixing of philosophies; and (c) site visits, focus groups, and individual interviews (See Table 18 and Figure 8 below). While the case study methods resonate only with academic evaluators, none of the quantitative methods cited in the evaluation literature resonate with any of the STEM evaluators interviewed.

Table 18. The Discourse of the Current and Future State of the STEM Evaluation Field: Evaluation Methodologies (STEM Evaluation Practitioner in Academic, n = 6; Private, n = 5; Academic/Private, n = 1; Public, n = 1)

Evaluation Literature	AEA STEM TIG: Abstracts			STEM Evaluation Practitioner: Interviews		
Methodologies	Academic	Private	Public	Academic	Private	Public
Quantitative Designs[1]						
Experimental Designs (RCTs)			✓			
Quasi-Experimental Designs (QEDs)	✓		✓			
Correlational Designs			✓			
Qualitative Methods[1]						
Case study	✓	✓		✓		
Status and Survey	✓	✓	✓	✓	✓	✓
Interpretive designs, Site Visits, Focus Groups, Individual interviews	✓	✓	✓	✓	✓	
Mixed Methods[1]						
Mixing of philosophies	✓	✓	✓	✓	✓	
Research designs	✓	✓	✓	✓	✓	
Data collection devices	✓	✓	✓	✓	✓	✓

1 - Lawrenz, F., & Huffman, D. (2006).

Figure 8. STEM Evaluation Methodologies Discussed in the Literature Compared to Perceptions of STEM Evaluators in Academic (1), Private (2), and Public (3) Sectors According to Their Interviews



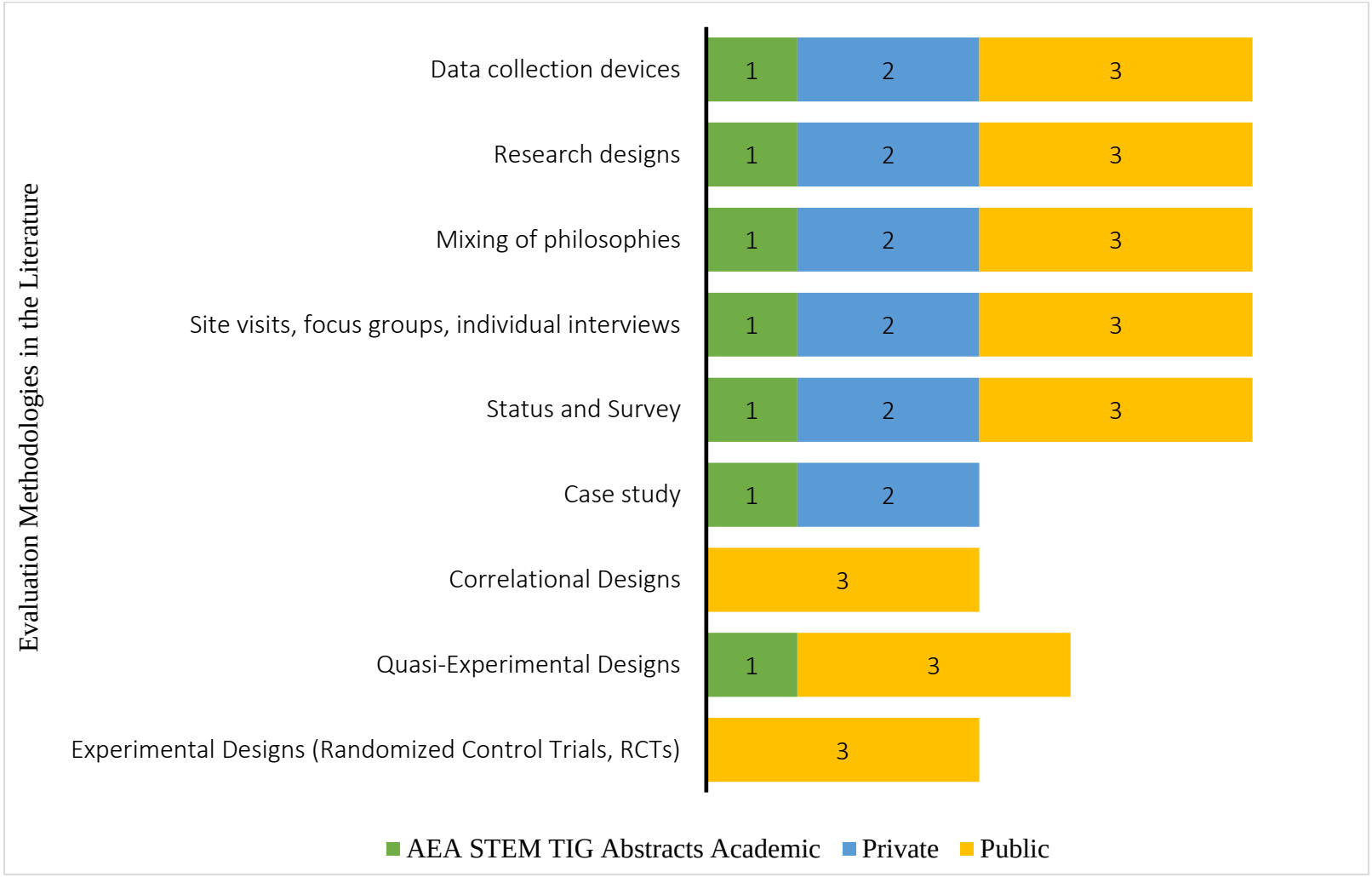
Evaluation Methodologies from the STEM Evaluation Literature versus AEA STEM TIG Abstracts

Evaluators in the STEM TIG abstracts discussed five evaluation methodologies that are like those found in the evaluation literature (Figure 9). The methodologies in the evaluation literature that resonate with all evaluators in the AEA STEM TIG abstracts include both mixed methods and qualitative methods. Specifically, in the mixed-methods category, all three components—philosophies, research designs, and data collection devices—resonated with AEA STEM TIG evaluators. Two out of three qualitative methods from the evaluation literature resonated with AEA STEM TIG evaluators, including (a) status and survey and (b) site visits, focus groups, and individual interviews.

Other findings in the data include the following:

1. Case Study Designs from the literature resonate well with academic and private sector evaluators at AEA STEM TIG.
2. Quasi-Experimental Designs from the literature resonate with academic and public sector evaluators.
3. Correlational Designs and Experimental Designs (Randomized Control Trials, RCTs) from the literature resonate with only the public sector evaluator

Figure 9. STEM Evaluation Methodologies Discussed in the Literature Compared to Perceptions of AEA STEM TIG Evaluators in Academic (1), Private (2), and Public (3) Sectors According to Their Abstracts



CHAPTER V: DISCUSSIONS AND CONCLUSIONS

This chapter includes a discussion of major findings as related to the literature on evaluation approaches and methodologies in STEM education contexts, a comparison of different approaches and methodologies, their major strengths, and gaps, and how and in what ways findings could influence the field of STEM education evaluation in its foreseeable future. Also included is a discussion of practitioners' thoughts and experiences about their work and their viewpoints on the prevailing Discourse in STEM Education Evaluation and the future direction of the field. This chapter concludes with a discussion of the limitations of the study, areas for future research, and a summary of final thoughts.

Specifically, in this chapter, I discuss five major findings related to the Discourse of STEM Education Evaluation.

As presented in Chapter Four, the Discourse surrounding STEM education evaluation typically focuses on lessons learned, program challenges and successes, program impact, and equity issues. Findings in this research showed discussion revolved around what evaluators learned from past experiences, the challenges faced in implementing STEM programs, the successes achieved and how they were achieved, the impact of the programs on students, educators, and different stakeholders and communities, and the equity issues that arise in access to STEM education and resources.

As the sample in this study shows, STEM evaluators encompass a diverse group of professionals from various fields and, due to their diverse backgrounds, they bring with them different perspectives, values, and beliefs about what constitutes the dominant Discourse of STEM Education Evaluation. Additionally, there is a growing call by evaluators in this study for a shift toward a plurality of methodologies in STEM evaluation. This call is driven by a desire

for more nuanced and context-sensitive evaluation approaches to determining what works and why.

By considering evaluation settings, program community, and contextual factors like intention, values, and equity issues, STEM evaluators can discuss and dialogue and gain a more complete understanding of the complexities of STEM education initiatives. In this research, there are five major takeaways for the Discourse of STEM Education Evaluation: STEM evaluators' current thinking, their position relative to the evaluation literature, and the future direction they imagined for the field:

1. STEM evaluators highly value professional development in all aspects of evaluation work and settings.
2. Collaborative evaluation approaches are widely popular among evaluators because of their emphasis on building trust and genuine partnerships across STEM education programs.
3. STEM evaluators continue to lean toward traditional methodologies that are characterized by representations of participants' experiences and data sense-making.
4. STEM evaluators are building a broader community of STEM participants and professionals with an equity focus and a sense of accountability rooted in self-socialization, leadership, and social justice.
5. The use of data in the era of new technology and new stakeholders is influencing the current and future direction of STEM education evaluation.

The Discourse of STEM Education Evaluation Focuses on Learning, Professional Growth, Program Improvement, and Consideration for Multiple Factors while Working in Evaluation Settings

Lessons Learned

The Discourse of STEM Education Evaluation in this research highlights several lessons learned by STEM evaluators. One of the primary lessons is the importance of professional development opportunities for STEM educators and program managers. The interview participants in this study emphasized the need for mentorship, career preparation, and work-life balance to support the success of STEM professionals. Additionally, interviewees emphasized the need for team-based projects and course activities that foster collaboration and critical thinking among students and program participants. Another lesson learned is the importance of program funding, which affects the sustainability and scalability of STEM programs.

The discourse highlights the need for program managers to diversify their funding sources, including grants, donations, and partnerships, to ensure the long-term success of the program. According to an analysis of abstracts from the American Evaluation Association (AEA) STEM evaluators, recruitment and retention of program participants are critical challenges for evaluators. Additionally, implementing large multisite projects and measuring collective impact are some of the significant challenges that evaluators face (Ståhl et al., 2016). Furthermore, evaluators are also tasked with developing effective data collection methodologies that can provide relevant and reliable data (Van Lange et al., 2013).

Program Improvement

The effectiveness of team-based projects, course activities, and program funding also form an integral part of the Discourse of STEM Education Evaluation. Evaluators assess

program success and effectiveness through a range of metrics, including the impact of team-based projects, course activities, and program funding. These metrics help evaluators understand the strengths and weaknesses of STEM programs, leading to improvements in program design and implementation (Ståhl, 2016).

Additionally, Discourses of STEM Education Evaluation include challenges associated with the recruitment and retention of program participants, implementing large multisite projects, data collection methodology, and collective impact measurement. STEM education evaluators in this research focused on professional development matters, program impact and effectiveness, and their work settings. By considering these factors, evaluators were able to improve their work and ensure that STEM education programs are effective and impactful.

Professional Growth

STEM evaluators also highlight the efforts to support their professional growth and success. The evaluators mention several professional development opportunities, including workshops, conferences, and training sessions. Additionally, the evaluators emphasize the importance of networking opportunities and community-building and organizing experiences in their line of work. According to the abstract review, evaluators must focus on mentorship, preparing for a career in science, and work-life balance (Durbin et al., 2019). This is critical to ensuring that evaluators are well-equipped to perform their duties and that they remain motivated and engaged in their work.

STEM Evaluators' Work Settings

Academic evaluators primarily work in STEM centers/colleges, private evaluators in K–12, higher education, museums, etc., and public evaluators in research institutions and summer academies. The following gives a closer look at evaluation settings by evaluators' sector:

Academic vs. Private and Public Sector Evaluators

Academic sector evaluators are typically found in higher education institutions such as colleges and universities, as well as K–12 school classrooms and STEM learning and research centers. Their focus is often on evaluating the effectiveness of educational programs, teaching methods, and learning outcomes, and they may also be involved in researching these topics. In contrast, private and public sector evaluators may work in a wider range of evaluation settings, such as museums and science centers, community learning centers, subways, out-of-home media, city parks, and social media.

Private vs. Academic and Public Sector Evaluators

Private sector evaluators work in a range of evaluation settings, like academic sector evaluators, but they are often found in the business sector and may work in companies that provide educational services, such as K–12 schools, higher education institutions, and research institutions. Additionally, private sector evaluators may work in museums and science centers, community learning centers, subways, out-of-home media, city parks, and on social media. This wider range of evaluation settings allows private sector evaluators to bring a different perspective and expertise to the field, compared to their academic and public sector counterparts.

Public vs. Academic and Private Sector Evaluators

Public sector evaluators are typically employed by government agencies and organizations, such as the Department of Education or the National Science Foundation. These evaluators focus on evaluating the effectiveness of STEM educational programs and initiatives funded by the government, as well as researching STEM education. Public sector evaluators also work in multisite or a variety of research institutions, summer academies, seminars, and workshops, and they use evaluation frameworks to guide their work. Compared to academic and

private sector evaluators, public sector evaluators have a unique perspective and knowledge about the policies and regulations that affect STEM education at the national and local levels.

In summary, each of these three sectors—academic, private, and public—brings a different perspective and expertise to the field of STEM education evaluation, and they each contribute to advancing the field in their way.

Examples of Academic, Private, and Public Sector Evaluators' Contributions to the Development of STEM Education

Academic Evaluators

Academic STEM evaluators are typically affiliated with universities and research institutions and are experts in their respective fields. Their expertise and experience in research and teaching help them bring a unique perspective to the evaluation of STEM education programs. One of the key contributions of academic evaluators is the development of new knowledge and understanding in their field. Through research studies, academic STEM evaluators can evaluate the effectiveness of various teaching and learning strategies and identify best practices to improve STEM education outcomes. They play a crucial role in evaluating the effectiveness of different teaching and learning strategies in STEM education. Through research, they can identify best practices and collaborate with others to improve student outcomes.

Another important aspect of the role of STEM evaluators is collaboration. STEM evaluators often work in interdisciplinary teams with stakeholders, including program managers, instructors, and students, to improve program outcomes. Collaborative approaches to evaluation can improve the quality and impact of STEM programs and initiatives (Huffman et al., 2006). The STEM program evaluation laboratory mentioned in the interviews with academic STEM evaluators is an example of a collaborative approach to STEM evaluation, as it engages

undergraduates in program evaluation and provides them with opportunities to pursue graduate degrees in the field while learning the rudiments of evaluation practice. Academic evaluators also contribute to STEM education by disseminating their research findings through academic publications. Their publications help to inform other researchers, policymakers, and educators about the latest research in STEM education. For example, the *Journal of Research in Science Teaching* is a leading academic journal that publishes research on science teaching and learning. The articles in this journal help to inform best practices in science education and contribute to the development of new knowledge in the field.

In conclusion, academic STEM evaluators play a crucial role in evaluating the effectiveness and impact of STEM programs and initiatives. The examples provided in this discussion illustrate the importance of equity, diversity, and inclusion considerations and collaboration in STEM evaluation. These findings agree with the STEM evaluation literature and highlight the need for evaluators to incorporate these considerations in their work.

Private Sector Evaluators

Private STEM evaluators play a crucial role in evaluating the effectiveness of STEM education programs, initiatives, and projects in various settings such as schools, museums, and other educational institutions. These evaluators are typically contracted by program sponsors to assess the impact of the program and provide recommendations for improvement. They use various research methods and data analysis techniques to gather evidence of program effectiveness, and they communicate their findings to program sponsors, educators, and other stakeholders.

Findings in the AEA STEM abstracts provided examples of how private STEM evaluators work to improve the quality of evaluation in different settings. For instance, the importance of collecting validity evidence within the science festival context highlights the need for stronger evaluation practices among the science festival community. Evaluators working in this setting can use a range of methods, such as surveys and interviews, to gather data on the impact of the festival on attendees and to establish the validity of these measures.

Another example mentioned in the AEA STEM abstracts is the integration of website development into STEM education training programs to facilitate evaluation. This approach is particularly useful for smaller professional development programs that have limited evaluation budgets. By embedding website development training into the broader professional development offering, evaluators can gather data on program impact and student learning without incurring additional costs. These examples are in line with the STEM evaluation literature, which emphasizes the importance of using multiple methods to collect data, establishing the validity of measures, and integrating evaluation into program design. For instance, Donaldson and Lipsey (2006) emphasize the need for evaluators to use a range of quantitative and qualitative methods to gather data and to establish the validity and reliability of their measures. Stufflebeam et al. (2003) argue that evaluation should be an integral part of program design and that evaluators should work closely with program designers and implementers to ensure that evaluation data are collected throughout the program cycle.

In summary, private STEM evaluators play an important role in evaluating the effectiveness of STEM education programs in various settings. The examples provided in the AEA STEM abstracts demonstrate the importance of collecting validity evidence and integrating

evaluation into program design, which is consistent with recommendations in the STEM evaluation literature.

Public Sector Evaluators

Public STEM evaluators play an important role in evaluating STEM programs and initiatives to determine their effectiveness and impact on various stakeholders. The primary roles of public STEM evaluators include assessing program outcomes, measuring program impact, and providing feedback to program staff and funders. They also play a crucial role in improving program design and implementation by identifying areas of strength and weakness and recommending changes to improve program effectiveness.

Public STEM evaluators' work can take place in various settings and contexts, such as academic institutions, government agencies, non-profit organizations, and private consulting firms. Evaluators may work independently or as part of a team and may employ various research methods, such as surveys, interviews, focus groups, observations, and document reviews. The AEA STEM evaluator abstract findings on the evaluation of the National Alliance for Partnerships in Equity (NAPE) STEM Equity Pipeline Programs provide a good example of how public STEM evaluators can use customized evaluation plans and partnerships to evaluate STEM programs. One specific example from the evaluation is the micro-messaging to Reach and Teach Every Student curriculum, which is a professional development program for secondary educators that aims to increase the participation, performance, and persistence of students in STEM, particularly those from underrepresented populations.

The evaluation aims to determine the extent to which the micro-messaging training changes the attitudes and behaviors of secondary STEM teachers, and if these changes in trained teachers produce changes in students' attitudes. The evaluation is ultimately focused on making

classrooms more equitable and increasing student success. The initial findings from teacher pre/post surveys, teacher observations, and student-level data were shared by the University's Center on Research and Evaluation to describe NAPE's impact on the evaluation aims.

This example agrees with STEM evaluation literature, which emphasizes the importance of evaluating STEM programs to determine their effectiveness and impact on various stakeholders, particularly those from underrepresented populations (Malyn-Smith, 2014). For example, the National Science Foundation emphasizes the importance of evaluating STEM programs to ensure they are meeting the needs of diverse populations and addressing equity and access issues (James & Singer, 2016). The American Evaluation Association also emphasizes the importance of cultural competence in STEM evaluation to ensure that evaluators are sensitive to the needs and perspectives of diverse stakeholders (Coryn et al., 2016).

In conclusion, public STEM evaluators play a crucial role in evaluating STEM programs and initiatives to determine their effectiveness and impact on various stakeholders. The example of the evaluation of the NAPE STEM Equity Pipeline Programs demonstrates how customized evaluation plans and partnerships can be used to evaluate STEM programs, particularly those focused on underrepresented populations. This example agrees with STEM evaluation literature, which emphasizes the importance of evaluating programs to ensure they are meeting the needs of diverse populations and addressing equity and access issues.

Finally, academic, private, and public sector evaluators each bring a unique perspective and expertise to STEM education evaluation. By understanding these differences, evaluation practitioners can design more effective and relevant STEM education programs that meet the needs of program communities, program managers, funders, the whole STEM enterprise, and the public.

STEM Education Evaluators Prioritize Collaboration, Cultural Responsiveness, Capacity Building, and Client-Centered, Utilization-Driven Evaluation Approaches

STEM education evaluators prioritize a range of evaluation approaches that are context-specific and stakeholder-centered. This includes collaborative evaluation approaches, culturally responsive evaluation, evaluation capacity building, client-centered evaluation, and utilization-focused evaluation approaches. The purpose of this section is to discuss these approaches in detail, drawing on relevant academic literature, including AEA STEM TIG abstracts and STEM evaluator interview participants.

Collaborative Evaluation Approaches

These are prioritized by STEM education evaluators because they emphasize the engagement of all stakeholders or partners, such as program staff, participants, funders, and policymakers. Collaborative evaluation approaches systematically engage these stakeholders in the evaluation process to facilitate the co-creation of a more complex understanding of programs (Coryn et al., 2016; Morell, 2010; Bickman & Rog, 2008). According to an AEA STEM evaluator abstract, collaborative evaluation approaches facilitate the co-creation of a more complex understanding of programs and describe the range of culturally relevant perspectives in program communities.

Additionally, the viewpoint of an academic sector interviewee in this study is that of a STEM education evaluator who values collaborative evaluation approaches. The interviewee believes that collaborative approaches help to facilitate stakeholder engagement, understanding of the program's complexity, and the creation of shared knowledge. Furthermore, the interviewee brings a perspective on evaluation in educational programs, with a focus on working with groups that have historically been marginalized in STEM disciplines. The interviewee is

concerned with ensuring that historically marginalized groups have a positive experience in STEM programs and that their voices are heard and represented. They believe that this is important because it can help program planners to develop programs that are more inclusive and effective, given the big push to bring in people who have been historically left out or pushed out in STEM disciplines.

Culturally Responsive Evaluation

Culturally responsive evaluation is another approach that has gained attention in STEM education evaluation. It is an approach that emphasizes the importance of incorporating cultural sensitivity into the evaluation process. This approach recognizes that cultural differences can influence the evaluation process and outcomes. According to a study by Rodriguez and colleagues (2014), culturally responsive evaluation can help evaluators identify potential barriers to learning and address them in the evaluation process. The study found that culturally responsive evaluation can improve the validity of evaluation findings and make them more applicable to diverse populations. Similarly, in the interviews with STEM evaluator participants, culturally responsive evaluation was seen as an important approach for addressing the needs of diverse student populations. One of the academic sector interviewees emphasized the importance of evaluating programs that are culturally relevant and responsive to students' needs. The participant recognized that it takes sensitivity and understanding of the program communities they are working with to do this. In the interview, the participant stated:

> And as STEM Educators, we're not just responsible for supporting them in the classroom. But if they don't know where their next meal is coming from, or they don't have housing, or they don't have basic social support, how are they expected to do something in the classroom?

The above quote highlights the need for educators to consider the social and emotional well-being of their students when developing and evaluating programs. To address this issue, the interviewee mentioned their efforts to develop an app in collaboration with an engineering department to connect students with resources that could provide psychosocial support. The participant stated:

> We are trying to develop an app. We work with engineering to develop an app that could connect students with resources that could meet these psychosocial supports and try to think about how we can use our evaluation tools to help with education like if we know as an evaluator this stuff is going on.

This quote shows the interviewee's recognition that as an evaluator, they have to be responsive to the social and emotional needs of students and use evaluation tools to support their academic success. Overall, this academic sector evaluator, like other interview participants, stressed the importance of being culturally responsive and recognizing the impact of social and emotional factors on students' academic success. They highlighted the need for educators and evaluators to consider the needs of the program communities they work with and develop tools to support their student's social and emotional well-being.

Evaluation Capacity Building (ECB)

Evaluation Capacity Building (ECB) refers to the processes, methods, and tools that support individuals and organizations in building their evaluation knowledge and skills (Wandersman et al., 2015). ECB is essential for strengthening the evaluation capacity of STEM education programs, which enables them to develop more effective interventions and allows stakeholders to acquire the necessary skills and knowledge to participate meaningfully in the evaluation process and promote ongoing program improvement (Labin, 2012; Wandersman et

al., 2015). In the AEA STEM TIG abstracts, ECB was frequently mentioned as a learning experience or tool to benefit STEM education program evaluation. Furthermore, in a study by Bozzo (2012), private-sector evaluators were found to be more likely to use ECB approaches than academic or government evaluators. The study suggests that private sector evaluators may be more focused on building capacity within the organizations they work with to improve the effectiveness of their programs.

The private sector evaluator in question emphasizes the importance of Evaluation Capacity Building (ECB) in building the capacity of organizations from cradle to mantle of leadership teams. The evaluator has experience working on a National Science Foundation (NSF) funded project where the goal was to build evaluation capacity among science festival partners across the US. The evaluator indicates that the project interviewed different science festival leaders and their evaluation staff to learn more about how they use evaluation and the skills they have learned through participating in the project.

Similarly, in the interviews with STEM evaluators, one of the private sector evaluators emphasized the importance of ECB in building the capacity of organizations from cradle to mantle of leadership teams. The evaluator gave an example of one of her projects and noted that the focus of that project was on building evaluation capacity. She described a range of activities such as helping people think about evaluation, finding value in evaluation, building skills to collect data, analyzing data, and disseminating or making meaning from the data. This finding is supported by academic literature, which suggests that ECB can enhance an organization's ability to plan, implement, and evaluate programs effectively (Preskill & Boyle, 2008). The evaluator's emphasis on ECB highlights the importance of developing the skills and knowledge necessary to conduct effective evaluations within an organization.

By building evaluation capacity, organizations can develop a culture of evaluation, which promotes continuous learning and improvement. This approach can lead to more effective decision-making, as organizations are better equipped to gather and analyze data to inform their actions (Huffman et al., 2008).

The emphasis on ECB also suggests that the evaluator believes that evaluation is an ongoing process, and organizations must continually develop their capacity to evaluate their work. Overall, the private sector evaluator views Evaluation Capacity Building as a critical component of building the capacity of organizations to use evaluation effectively. Through their experience with the NSF project, the evaluator recognizes that ECB can take many forms and can involve developing a range of skills to support data collection, analysis, and dissemination. By emphasizing the importance of building evaluation capacity, the private sector evaluator highlights the need for organizations to prioritize evaluation as a tool for continuous improvement and decision-making.

Client-Centered Evaluation

Client-Centered Evaluation is an evaluation approach that focuses on the needs and interests of the stakeholders or clients involved in the evaluation process. This approach recognizes the importance of involving stakeholders in the evaluation process and aims to provide feedback and recommendations that are relevant and meaningful to them. In a study by Patton (2011), client-centered evaluation was found to be an effective approach for engaging stakeholders in the evaluation process and for ensuring that the evaluation was responsive to their needs. The study found that this approach can help to build trust between the evaluator and the clients and can lead to more useful evaluation findings. Overall, the AEA STEM TIG abstracts noted that client-centered evaluation is critical in STEM education because it helps to

ensure that programs meet the needs and priorities of the clients, ensuring that the evaluation is relevant, useful, and actionable (Rauschecker, 2011; Ardoin & Bowers, 2020). This approach gives stakeholders a voice in the evaluation process, ensuring that their perspectives are considered (Donaldson & Lipsey, 2006). STEM evaluators in the interview added strength to the Client-Centered approach in shining light on program participants and surfacing their voices. According to the findings in this study, a private sector evaluator, Ugochukwu, believed that this approach is valuable in highlighting the experiences and perspectives of program participants. Ugochukwu emphasized the need to balance quantitative data with qualitative insights from program participants, stating:

> I hear those voices, whereas many people, typically the ones in charge of budgets, like to see that it was, you know, 90%, or they want to see it was a two standard deviation improvement. And, you know, balancing both of those, I think, is a requirement to understand what's going on at a meaningful level, and be able to make meaningful changes with that data.

This perspective aligns with the existing academic literature on program evaluation, which suggests that a Client-Centered approach can lead to more meaningful and accurate evaluations (Fetterman & Wandersman, 2005). This approach emphasizes the importance of involving program participants in the evaluation process and valuing their input and feedback. By doing so, evaluators can gain a more comprehensive understanding of program outcomes and make more informed decisions about how to improve the program.

In addition, the interviewees' emphasis on balancing quantitative data with qualitative insights echoes the literature on mixed-methods evaluations. As Creswell and Plano Clark (2017) note, mixed-methods evaluations can provide a more complete picture of program outcomes by

combining the strengths of both quantitative and qualitative data. By integrating both types of data, evaluators can gain a more nuanced understanding of program outcomes and use this information to make meaningful improvements. Overall, the interviewees' thoughts and impressions on the importance of a Client-Centered approach and the need to balance quantitative and qualitative data are supported by the academic literature on program evaluation. By valuing program participants' voices and using a mixed-methods approach, evaluators can gain a more comprehensive understanding of program outcomes and make more informed decisions about program improvement.

Utilization-Driven Evaluation Approaches

Also known as Utilization-Focused Evaluation (UFE), this is an approach that emphasizes working closely with program stakeholders to identify evaluation questions, design data collection methods, and develop evaluation findings that are relevant and useful for them (Alkin, 2004; Newcomer et al., 2015). This approach is critical for ensuring that evaluation findings are used to improve programs and achieve their intended outcomes. The findings of this study align with the literature regarding the importance of UFE for ensuring that evaluation findings are used to improve programs and achieve their intended outcomes. According to an AEA STEM evaluator abstract, the UFE approach is particularly relevant to STEM education programs, as it ensures that the evaluation results are used to improve the program's impact on students' learning and career success (Ardoin & Bowers, 2020).

Another AEA STEM evaluator abstract stated its presentation will, "describe the UFE method used within one out-of-school time STEM setting to develop and refine program goals, examine program implementation and efficacy, provide mid-course feedback on program delivery, and reflect on program strategies." In addition, the UFE approach was highlighted by

some of the STEM evaluators interviewed in this study, who emphasized the importance of making sure that evaluations are not just done for the sake of the evaluation, but are useful and actionable (Askew et al., 2012). Similarly, STEM evaluators interviewed in the study by Finelli (2012) identified UFE as an important evaluation approach in the STEM education context. The participants in that study emphasized the need for evaluators to design evaluations that are relevant and useful for decision-making and to ensure that evaluation findings are communicated in a clear and accessible manner.

In summary, STEM education evaluators use various evaluation approaches to ensure they meet the needs of the program and its stakeholders. Collaborative Evaluation Approaches, Culturally Responsive Evaluation, Evaluation Capacity Building, Client-Centered, and Utilization-Driven evaluation approaches (Evaluation Frameworks) are critical approaches in STEM education evaluation. Evaluation questions drive the choice of the evaluation approach used by STEM education evaluators, and there are differences in the perspectives of evaluators in the academic, private, and public sectors. It is essential to understand these differences and use them to develop more effective STEM education evaluation approaches.

Mixing of Philosophies Methodologies Dominates the Conversion, and Followed by Traditional, Zigzag, and Scientific Methodologies

STEM Evaluation Methodologies

STEM education evaluation is a complex process that requires careful consideration of a range of distinct philosophies that help articulate the strategies deployed in qualitative, quantitative, or mixed-methods research. According to the AEA STEM evaluator abstracts, the main evaluation methodologies identified include traditional methodologies, Zigzag methodologies, mixed methodologies, and scientific methodologies. Similar methodologies exist

with STEM evaluators, except there was no mention of scientific methodologies during the private sector participant interviews. Each of the methodologies recorded has unique characteristics and assumptions that inform their application in STEM education evaluation.

The Traditional Methodologies

These are characterized by adhering to industry-validated and standardized ways of knowing and being, which are designed to represent experiences and sense-making in a more standardized and consistent manner. The AEA STEM evaluator abstracts surface this approach with a heavy focus on quantitative data, and evaluators who describe traditional methodologies in their presentation aim to measure program effectiveness using predetermined, standardized metrics. This justification of methods of evaluation aligns with the positivist philosophical perspective, which assumes that there is a single objective reality that can be measured through standardized methodologies (Lincoln & Guba, 1985).

Zigzag Methodologies

The Zigzag Methodologies are opposed to normative components of sense-making and pursue non-linear, meandering paths and connect the articulation of findings with theoretical orientations that drive the choice of methodologies employed in evaluation. These methodologies show up in the AEA STEM evaluator abstracts intending to account for the complexity of the context in which STEM programs are evaluated, and therefore take a more contextual and interpretive approach to evaluations. This approach aligns with the constructivist philosophical perspective, which assumes that reality is socially constructed, and that meaning is created through the interpretation of experiences (Lincoln & Guba, 1985).

Mixed Methodologies

The mixed methodologies are widely used in STEM education evaluation and entail the philosophical assumptions underlying the application of qualitative and quantitative methodologies, and the incorporation of both approaches in tandem so that the overall strength of a study is greater than either of them (Creswell & Plano Clark, 2017). This approach combines the strengths of both positivist and constructivist perspectives and allows for a more comprehensive and nuanced understanding of STEM education programs.

Scientific Methodologies

The scientific methodologies establish assumptions about the regularity of causation and focus interrogation of inquiry as to what (happened) rather than to the why (justification) it happened. The scientific methodologies show up in the AEA STEM evaluator abstracts, and its overall objective aligns with the positivist philosophical perspective that is typically used in the natural sciences, where causality and regularity are assumed to be defining features of the natural world.

In the Discourse of STEM Education Evaluation, evaluators pay attention to issues of identity, status, and power in their program domains. The STEM education evaluators use various methodologies, but the most cited methodologies emerging from interview discussions with STEM evaluators are mixed methodologies, followed by collaborative methodologies (AEA STEM Evaluator Interview Participants Dialogue, n.d.).

Evaluators in the STEM fields provide the most justifications for Data Collection Devices, and Status and Surveys for evaluation methodologies. Research findings revealed that mixed methods (data collection devices), and qualitative methods (status and survey) were the two most frequently discussed evaluation methodologies during interviews with STEM

evaluators. For example, an academic evaluator during an interview opined those mixed methods are valuable for the future of STEM program evaluation given the social justice issues the nation is currently facing (AEA STEM Evaluator Interview Participants Dialogue, n.d.). This highlights the importance of using a mixed-method approach that considers both the quantitative and qualitative aspects of STEM programs, as well as the broader social and political contexts in which they operate.

In conclusion, the Discourse of STEM Education Evaluation is characterized by a range of philosophies and methodologies, including Traditional Methodologies, Zigzag Methodologies, Mixed-Methodologies, and Scientific Methodologies. While evaluators use various methodologies, the following are earmarks of most STEM education evaluations.

STEM Education Evaluators Value Equity, Professional Development, and Community Organizing, and Evolve Through Self-Socialization, Leadership, and Social Justice

STEM education has become an increasingly important topic, with a growing focus on equity, professional development, and community organizing. Evaluators of STEM education programs and projects are discussing these issues in various settings, including NSF- and NIH-funded STEM programs in colleges and universities, STEM learning, research laboratories, after-school STEM programs, and professional development workshops.

Professional Development

Professional development is a critical aspect of STEM education evaluation, and it is a common evaluation setting for all three including the private, public, and academic sectors. The focus of evaluation conversations varies depending on the sector. Academia often caters to STEM students, teachers, faculty, program managers, and funders; the private sector focuses on youth, teachers, evaluators, researchers, and designers, and public practitioners pay close

attention to the youth, staff, program managers, administrators, administrators, and program funding agencies (AEA STEM evaluator abstracts).

Academic sector evaluators frequently discuss professional development concerning issues of data collection challenges, training of STEM teachers, and learning opportunities. They also design frameworks, methods, tools, and tips for evaluating K–12 STEM educators and professionals, evaluate the inter-relationship between teacher professional development and student achievement, and discuss challenges in transforming their curricula and classroom practice (Waite & McDonald, 2019).

Equity, Community Organizing, and Evolving Through Self-Socialization, Leadership, and Social Justice

Alongside dimensions of diversity and inclusions, equity constitutes a critical issue that evaluators in STEM education frequently discuss. According to the STEM evaluator interview participants, the equity should be the responsibility of everyone involved in the STEM education ecosystem to achieve the broader goals of NSF Broadening Participation. An academic sector interview participant emphasized the importance of STEM evaluators being allies in the equity process. According to the participant, being an ally in the equity process requires evaluators to share power with program communities and engage in advocacy work. The participant argues that evaluators need to challenge STEM education teams to be more attentive to issues of diversity and broaden their understanding of diversity beyond just race and ethnicity.

The participant suggests that by being critical friends, evaluators can play a vital role in creating a shift in the STEM education field. By advocating for and educating STEM education teams on issues of equity, evaluators can encourage them to focus on these issues throughout their projects. The participant also suggests that evaluators need to broaden STEM education

teams' understanding of diversity to include leadership, content, and other aspects. The participant's viewpoint underscores the importance of evaluators actively engaging with STEM education teams to promote equity and inclusivity. By advocating for and challenging STEM education teams, evaluators can help broaden their understanding of diversity and encourage them to focus on issues of equity throughout their projects. This can ultimately lead to a more equitable and inclusive STEM education field.

Community Building

A dimension of community building and organizing practices that showed up during interviews with STEM evaluators is mentoring, where participants actively seek and engage in professional relationship building with a more experienced practitioner/mentor or less experienced colleague or mentee. This enables the sharing of knowledge and skills, building trust, and creating a supportive environment for all stakeholders (Grant, et al., 2016).

For programs and projects funded by the Department of Education or NSF funded, evaluators can explore student-level outcomes to see if the intervention at the classroom level is impacting all students or if one demographic is left behind. Technology plays a crucial role in evaluating STEM education programs and projects, and for evaluators in the public sector, it is not just a tool but an integral part of the evaluation practice. It enables conversation starters, facilitates, and smoothens issues during meetings, and is key to data integration in central databases. Lessons learned in using technology to better inform practice and move the conversation around EDI forward are paramount to the Discourse of STEM Education Evaluation (Groff, 2013).

In conclusion, the Discourses of STEM Education Evaluation center around equity, professional development, and community organizing, which evolve through self-socialization,

leadership, and social justice. Evaluators must be allies in the process of promoting EDI, and mentoring is a critical component of building supportive and productive relationships in STEM education evaluation. Technology plays a crucial role in facilitating evaluation conversations and data integration, particularly in the public sector, where evaluation budgets are more expensive. Overall, STEM education evaluation must be focused on promoting inclusive practices that benefit all learners and stakeholders.

The State of STEM Education Evaluation is Influenced by Data Utilization in the Age of New Technology and New Stakeholders

STEM education evaluation has undergone significant transformations in recent times, as a result of the increasing role of new technology, data utilization, and new stakeholders and funders. The interplay of these factors has impacted the current and future states of STEM evaluation along various dimensions, including evaluation use, hybrid evaluation, virtual forms of data collection, face-to-face work, and evaluation approaches and methodologies.

Evaluation Use

Evaluation use refers to the extent to which evaluation findings are utilized by stakeholders to inform decision-making processes. The rise of new technologies has increased the potential for data utilization, allowing for more effective and efficient evaluation use. For example, the use of online surveys and virtual focus groups has become more common, providing evaluators with access to larger and more diverse samples of stakeholders (Boyce, 2017). This increased reach has the potential to improve the utilization of evaluation findings, as more stakeholders are involved in the evaluation process.

Hybrid Evaluation

The emergence of hybrid evaluation is another area where the impact of new technology is evident. Hybrid evaluation combines elements of traditional, face-to-face evaluation with virtual evaluation methods. This approach is effective in providing a more comprehensive evaluation of STEM education programs, as it allows for a wider range of data collection methods to be employed (Kivunja & Kuyini, 2017). The use of hybrid evaluation has the potential to improve the accuracy and reliability of evaluation findings.

Virtual Forms of Data Collection

Virtual forms of data collection have become increasingly popular in STEM education evaluation. These methods include online surveys, virtual focus groups, and virtual interviews. Virtual data collection methods offer several advantages, including greater convenience, increased accessibility, and reduced costs (Boyce, 2017). The use of virtual data collection methods has the potential to improve the quality and quantity of data collected in STEM education evaluations.

Face-to-Face Work

Despite the increasing use of virtual data collection methods, face-to-face work remains an important aspect of STEM education evaluation. Face-to-face interactions can provide a level of depth and richness that is difficult to achieve through virtual interactions alone. For example, in-person interviews and focus groups can provide evaluators with a greater understanding of stakeholders' experiences and perspectives (Kivunja & Kuyini, 2017). The use of face-to-face work in STEM education evaluation is likely to remain important, particularly in situations where nuanced or complex data are required.

Evaluation Approaches and Methodologies

The impact of new technology, data utilization, and new stakeholders and funders has also influenced evaluation approaches and methodologies in STEM education evaluation. For example, the use of mixed-methods evaluation has become more common, allowing for the integration of both qualitative and quantitative data (Kivunja & Kuyini, 2017). The use of big data analytics has also become more prevalent in STEM education evaluation, providing evaluators with the ability to analyze large amounts of data more efficiently and effectively (Boyce, 2017).

Thoughts and Impressions of STEM Education Evaluators

STEM education evaluators have varied thoughts and impressions regarding the impact of new technology, data utilization, and new stakeholders and funders on STEM education evaluation. Some evaluators are optimistic about the potential of new technology to improve the accuracy and reliability of evaluation findings, while others are more cautious, citing concerns about the potential for bias and the need to ensure that evaluation methods are appropriate for the context in which they are being used (Boyce, 2017). There is also concern about the potential for data overload, and the need to ensure that evaluation findings are presented in a way that is accessible and understandable for stakeholders (Kivunja & Kuyini, 2017).

In conclusion, the current and future states of STEM education evaluation are influenced by the role of new technology, data utilization, and new stakeholders and funders who are changing the Discourse of STEM Evaluation. Stakeholders are interested in seeing evidence of program impact and return on investment (ROI). They want to know that their investments in STEM programs are yielding measurable outcomes and making a difference in the lives of students and the programming community. Therefore, evaluators must demonstrate the value of

their work by providing evidence-based evaluations that can demonstrate program effectiveness and impact (Newcomer et al., 2015).

Reflection: Why the AEA STEM TIG Accepted the Fewest Abstracts in 2014?

When compared to subsequent years, the lowest number of abstracts accepted at the 2014 AEA conference in the (see Table 6) area of STEM education evaluation could be attributed to various factors. The factors include the recent establishment of the STEM topical interest group (TIG). According to Stevahn et al. (2020), AEA established the STEM topical interest group in 2014 as a response to the growing interest in the evaluation of STEM education programs. As a new group, it was expected to take some time to gain momentum and attract submissions. Chomphuphra, et al. (2019) also suggested that the lack of familiarity with the new interest group and its focus on STEM education evaluation could have contributed to the low number of submissions. This view is consistent with the observations of Burns (2011), who noted that it can take time for a new group to gain visibility and build a reputation in the evaluation community. Burns (2011) further argued that the establishment of a new interest group could also affect the review process, as reviewers may be less familiar with the new group's focus and evaluation criteria.

Moreover, according to Rizzolo, et al. (2016), it is not uncommon for conferences to experience fluctuations in the number of submissions, with some years recording more or fewer submissions than others. Therefore, the low number of abstracts accepted in 2014 may have been a result of chance or other factors beyond the establishment of the new STEM interest group. In summary, the low number of abstracts accepted at the 2014 AEA conference in the area of STEM education evaluation was likely due, at least in part, to the recent establishment of the STEM topical interest group. Establishing a new group often takes time to build momentum and

attract submissions, and there may have been limited familiarity with the group and its focus on STEM education evaluation, leading to fewer submissions in the first year. However, it is worth noting that fluctuations in the number of submissions are not uncommon at conferences, and other factors beyond the new interest group may have also contributed to the low number of abstracts accepted in 2014.

Implications of Findings

STEM education evaluation has become increasingly important as the need for a skilled workforce in STEM fields continues to grow (Burke, 2019). In this study, the Discourse on STEM Education Evaluation has been characterized by several important themes that have implications for the evaluation of STEM education programs. First, the Discourse on STEM Education Evaluation focuses on lessons learned, professional growth, and program improvement, considering multiple factors while working in different evaluation settings. STEM education evaluators recognize the importance of evaluating student learning outcomes as well as the professional development of educators in the program. They also consider a range of factors that affect the implementation and success of STEM education programs, such as curriculum design, classroom practices, and community engagement (Irwanto, et al., 2022). This finding implies that STEM education evaluation should be viewed as a complex and dynamic process that requires evaluators to consider a range of factors and stakeholders.

Second, STEM education evaluators prioritize collaboration, cultural sensitivity, capacity building, and client-centered and utilization-driven evaluation approaches. This finding suggests that STEM education evaluation is best done through a collaborative and culturally sensitive approach that involves all stakeholders in the evaluation process. STEM education evaluators should prioritize capacity building and focus on developing evaluation

strategies that are client-centered and utilization-driven, meaning that the evaluation results should be useful to stakeholders for making decisions about the program's improvement (Shade et al., 2021). This finding implies that evaluators must be skilled at engaging with stakeholders, and evaluation results should be presented in a way that is accessible and relevant to different audiences.

Third, the mixing of methodologies dominates the conversation, followed by traditional, Zigzag, and scientific methodologies. STEM education evaluators often employ a range of methodologies in their evaluations, including both quantitative and qualitative methods. This finding implies that evaluators should be familiar with a range of methodologies and be able to select the most appropriate method(s) for the specific evaluation context. Additionally, evaluators should be skilled at combining methods and interpreting results from multiple sources. Fourth, the states of STEM education evaluation are influenced by data utilization in the age of new technology and new stakeholders. With the proliferation of new technologies and the increasing involvement of diverse stakeholders in STEM education programs, the collection and utilization of data have become an increasingly important component of the evaluation process. Evaluators must be able to collect, analyze, and communicate data effectively to ensure that program stakeholders can make informed decisions about program design and implementation (Bryk et al., 2015).

Finally, STEM education evaluators value equity, professional development, and community organizing and evolve through self-socialization, leadership, and social justice. This finding suggests that STEM education evaluators must prioritize equity and consider how the evaluation process can help address disparities in STEM education access and outcomes. Additionally,

evaluators should focus on their own professional development and leadership skills to improve the quality and impact of their evaluations.

In sum, evaluators should consider how the evaluation process can support community organizing and social justice efforts (Staples, 2012). This finding implies that STEM education evaluators must be committed to social justice and equity, and they should strive to be leaders in the field.

In conclusion, the Discourse on STEM Education Evaluation is complex and multifaceted, with important implications for evaluators. STEM education evaluators must prioritize collaboration, cultural sensitivity, capacity building, and client-centered and utilization-driven evaluation approaches and be skilled at using a range of evaluation methodologies. Furthermore, the increasing utilization of data in the age of new technology and new stakeholders and the focus on equity, professional development, and community organizing are key factors that evaluators must consider in their work. By embracing these themes and engaging in ongoing professional development, STEM education evaluators can help to promote the success of STEM education programs and support the development of a skilled and diverse workforce in STEM fields.

Future Research

The Discourse of STEM Education Evaluation has become increasingly important in recent years, particularly in terms of understanding how STEM education programs are evaluated and how they can be improved. The annual American Evaluation Association (AEA) conference is one platform where evaluators from diverse backgrounds come together to share their insights on this topic. In this context, several possible future research questions can be explored. These include the following:

1. What are the current challenges that STEM education evaluators face when conducting evaluation, and how can these challenges be addressed through the use of innovative evaluation approaches and methodologies? What best practices have emerged in the field that could be replicated by other evaluators?
2. In what ways do professional development opportunities for STEM evaluators impact the quality of evaluation work and the outcomes of STEM education programs?
3. How do collaborative evaluation approaches contribute to building trust and genuine partnerships across STEM education programs, and what are the most effective ways to implement these approaches?
4. How can STEM evaluators best promote equity focus and accountability rooted in self-socialization, leadership, and social justice within STEM education programs, and what are the most effective strategies for building a broader community of STEM participants and professionals?
5. In what ways are new technologies and stakeholders impacting the use of data in STEM education evaluation, and how can STEM evaluators best adapt to these changes to ensure the most effective use of data in evaluating STEM education programs?

When considering these research questions, it is important to consider the perspectives and experiences of STEM education evaluators who are actively engaged in the field. The Center for Culturally Responsive Evaluation and Assessment (CREA) and the American Educational Research Association (AERA) are two organizations that provide valuable resources and guidance for evaluators in the field of STEM education. By reviewing the abstracts from these and other national/international evaluation conferences, researchers can gain a better

understanding of the current state of the field and identify areas where additional research is needed. The following research topics or questions could be pursued:

6. How do the Discourses of STEM education evaluation differ across different conferences (AEA, CREA, and AERA) and what are the implications of these differences for evaluating STEM education programs?
7. What are the most common evaluation methods and tools being used in STEM education evaluation across different conferences (AEA, CREA, and AERA), and how effective are these methods in evaluating STEM education programs?
8. What are the gaps and opportunities for future STEM education evaluation research across different conferences (AEA, CREA, and AERA), and how can evaluation practitioners and researchers work collaboratively to address these gaps and leverage these opportunities?

Expanding the research questions to include other national and international evaluation conferences could involve exploring the following future research questions:

9. How do the Discourses of STEM Education Evaluation differ across national and international evaluation conferences, and what are the implications of these differences for evaluating STEM education programs?
10. How can evaluation practitioners and researchers from different national and international evaluation conferences collaborate to develop more effective evaluation approaches for STEM education programs, and what are the key opportunities and challenges for this collaboration?

Essentially, future research on the Discourse of STEM Education Evaluation can help to improve evaluation approaches and methodologies, taking into account the thoughts and

impressions of STEM education evaluators. Future research can also address current and emerging challenges in the field, such as the need to incorporate culturally responsive evaluation approaches and methodologies into STEM education evaluation efforts. The review of AEA, CREA, AERA and other evaluation conference abstracts can help researchers identify current trends, challenges, and best practices in the field of STEM education evaluation and guide the development of more effective evaluation strategies to improve the quality of STEM education programs.

Concluding Thoughts: Key Take-Aways from this Research

This study on the Discourse of STEM Education Evaluation concludes with several important takeaways on the current state and future direction of STEM education evaluation.

The first takeaway is that STEM evaluators must engage in continuous professional development to enhance their technical and soft skills, including communication, collaboration, and leadership. This is particularly critical in areas such as STEM content knowledge, research methods, and evaluation tools and approaches. Ongoing professional development is essential for ensuring that STEM evaluators possess the necessary skills and knowledge to carry out thorough and effective evaluations. A recent study by Council & National Academies of Sciences, Engineering, and Medicine (2016) emphasized the importance of investing in professional development, given the gaps in approaches and methodologies used in STEM education evaluation. Furthermore, the Covid-19 pandemic highlights the need for STEM evaluators to adapt to new evaluation practices continually. Therefore, investing in professional development is essential for advancing and enhancing STEM education evaluation.

The second takeaway is that collaborative evaluation approaches are popular among evaluators and align with participatory and empowerment evaluation principles, which

emphasize stakeholder involvement in the evaluation process. Involving stakeholders in the evaluation process ensures that evaluation questions, methods, and results are aligned with their needs and perspectives. These findings contribute to the discussion of STEM education program evaluation and approach utilized by practitioners, which is significant as there is a paucity of knowledge in the evaluation literature regarding the Discourses surrounding these approaches and methodologies.

The third takeaway is that STEM education evaluators commonly use traditional and mixed methodologies, such as surveys, interviews, and focus groups, to collect data on participants' experiences and perceptions of STEM education programs. These types of methodologies provide a detailed analysis of the data, offering insights into the complex factors that influence STEM education outcomes. These findings add to the Discourse surrounding STEM education program evaluation, addressing a lack of knowledge regarding the approaches and methodologies utilized in STEM education evaluation.

The fourth is that this study emphasizes the evolving Discourse around STEM education evaluation, suggesting the importance of using new technologies and involving new stakeholders to improve data collection and analysis. The involvement of parents and community members can provide valuable insights, and the use of online surveys and data visualization tools can enhance the effectiveness of evaluation practices. These findings contribute to the discussion on the future direction of STEM education program evaluation and highlight the need to adapt to new technology and stakeholder involvement to enhance evaluation practices.

Finally, the Discourse on STEM Education Evaluation is at a critical juncture considering the unprecedented issues and challenges facing the nation in the wake of Covid-19 and post-racial reckonings. This research showed that STEM evaluators value professional development,

collaborative evaluation approaches, traditional and mixed methodologies, data-driven decision-making, and building a broader community of STEM participants and professionals with an equity focus. By embracing these takeaways, evaluators can produce more useful and meaningful evaluations that contribute to the improvement of STEM education programs and outcomes.

REFERENCES

Abdulwahed, M., Jaworski, B., & Crawford, A. (2012). Innovative approaches to teaching mathematics in higher education: a review and critique.

Adamski, T., Kline, B., & Tyrrell, T. (2006). FEMA reorganization and the response to hurricane disaster relief. *Perspectives in Public Affairs*, *3*, 3–36.

Alemdar, M., Cappelli, C. J., Criswell, B. A., & Rushton, G. T. (2018). Evaluation of a Noyce program: Development of teacher leaders in STEM education. *Evaluation and Program Planning*, *71*, 1–11.

Alkin, M. C. (Ed.). (2004). Evaluation roots: Tracing theorists' views and influences. Sage.

Allen, S., & Peterman, K. (2019). Evaluating informal STEM education: Issues and challenges in context. New Directions for Evaluation, 2019(161), 17-33.

Allen-Ramdial, S. A. A., & Campbell, A. G. (2014). Reimagining the pipeline: Advancing STEM diversity, persistence, and success. BioScience, 64(7), 612–618.

Anon. (2013). Raising the bar: reviewing STEM education in America: hearing before the Subcommittee on Early Childhood, Elementary, and Secondary Education. In *Committee on Education and the Workforce, U.S. House of Representatives, One Hundred Thirteenth Congress, first session*. Washington, DC: U. S. Government Printing Office.

Ardoin, N. M., & Bowers, A. W. (2020). Early childhood environmental education: A systematic review of the research literature. Educational Research Review, 31, 100353.

Ashby, C. M. (2006). Testimony before the Committee on Education and the Workforce, House of Representatives. Higher education: Science, technology, engineering, and mathematics trends and the role of federal programs.

Askew, K., Beverly, M. G., & Jay, M. L. (2012). Aligning collaborative and culturally responsive evaluation approaches. Evaluation and Program Planning, 35(4), 552-557.

Babbie, E. (2010). The practice of social research (13th ed.). Belmont, CA: Wadsworth Cengage Learning.

Bailey, R. A., Benton, T., & Challinor, A. J. (2015). Extreme weather and resilience of the global food system.

Barakos, L., Lujan, V., & Strang, C. (2012). Science, Technology, Engineering, Mathematics (STEM): Catalyzing Change Amid the Confusion. *Center on Instruction*.

Beard, R. L., Fetterman, D. J., Wu, B., & Bryant, L. (2009). The two voices of Alzheimer's: attitudes toward brain health by diagnosed individuals and support persons. *The Gerontologist*, *49*(S1), S40–S49.

Berch, D. B., Cordova, D. I., Earle, J. M., Humes, B. A., Olds, B. M., Levitt, J., & Sladek, M. F. (2006). Review and Appraisal of the Federal Investment in STEM Education Research. In. Washington, DC: Executive Office of The President, National Science, and Technology Council.

Bickman, L., & Rog, D. J. (Eds.). (2008). The SAGE handbook of applied social research methods. Sage publications.

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research*, *10*(2), 141–163.

Blackley, S., & Howell, J. (2015). A STEM narrative: 15 years in the making. *Australian Journal of Teacher Education*, *40*(7), 102–112.

Bowen, G. A. (2009). Document analysis as a qualitative research method. Qualitative Research Journal, 9(2), 27–40.

Boyce, A. S. (2017). Lessons learned using a values-engaged approach to attend to culture, diversity, and equity in a STEM program evaluation. *Evaluation and Program Planning, 64*, 33–43.

Bozzo, S. L. (2002). Evaluation capacity building in the voluntary/nonprofit sector. Canadian Journal of Program Evaluation, 17(3), 75-92.

Breiner, J. M., Harkness, S. S., Johnson, C. C., & Koehler, C. M. (2012). What is STEM? A discussion about conceptions of STEM in education and partnerships. School Science and Mathematics, 112(1), 3–11.

Brinkmann, S. (2014). Unstructured and semi-structured interviewing. *The Oxford handbook of qualitative research*, 277–299.

Brown Urban, J., & Trochim, W. (2009). The Role of Evaluation in Research—Practice Integration Working Toward the ‘‘Golden Spike’’. *American Journal of Evaluation, 30*(4), 538–553.

Brown, R., Brown, J., Reardon, K., & Merrill, C. (2011). Understanding STEM: Current perceptions. Technology and engineering teacher, 70 (6), 5–9. *View at*.

Browne, K. (2005). Snowball sampling: using social networks to research non-heterosexual women. *International journal of social research methodology, 8*(1), 47–60.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). Learning to improve: How America’s schools can get better at getting better. Harvard Education Press.

Burbules, N. C. (2004). Ways of thinking about educational quality. *Educational researcher, 33*(6), 4–10.

Burke, A. (2019). Science and engineering labor force. National Center for Science and Engineering Statistics. https://ncses. nsf. gov/pubs/nsb20198/introduction.

Burns, M. (2011). Distance education for teacher training: Modes, models, and methods.

Cannady, M. A., Greenwald, E., & Harris, K. N. (2014). Problematizing the STEM pipeline metaphor: Is the STEM pipeline metaphor serving our students and the STEM workforce? *Science Education*, *98*(3), 443–460.

Caracelli, V. J., & Greene, J. C. (1997). Crafting mixed-method evaluation designs. *New Directions for Evaluation*, *74*, 19–32.

Carter, S. M., & Little, M. (2007). Justifying knowledge, justifying method, taking action: Epistemologies, methodologies, and methods in qualitative research. *Qualitative health research*, *17*(10), 1316–1328.

Charmaz, K. (2014). *Constructing grounded theory*. sage.

Charmaz, K. (2017). The power of constructivist grounded theory for critical inquiry. *Qualitative Inquiry*, *23*(1), 34–45.

Charmaz, K., & Belgrave, L. (2012). Qualitative interviewing and grounded theory analysis. *The SAGE handbook of interview research: The complexity of the craft*, *2*, 347–365.

Chesky, N. Z., & Wolfmeyer, M. R. (2015). *Philosophy of STEM education: A critical investigation*. Springer.

Chomphuphra, P., Chaipidech, P., & Yuenyong, C. (2019, October). Trends and research issues of STEM education: A review of academic publications from 2007 to 2017. In Journal of Physics: Conference Series (Vol. 1340, No. 1, p. 012069). IOP Publishing.

Committee on Equal Opportunities in Science and Engineering (CEOSE). (2004). Broadening participation in America's Science and Engineering Workforce. In *The 1994–2003 Decennial and 2004 Biennial Reports to Congress*: Committee on Equal Opportunities in Science and Engineering (CEOSE).

Compton, D., Baizerman, M., Preskill, H., Rieker, P., & Miner, K. (2001). Developing evaluation capacity while improving evaluation training in public health: The American Cancer Society's Collaborative Evaluation Fellows Project. *Evaluation and Program Planning*, *24*(1), 33–40.

Coryn, C. L., Ozeki, S., Wilson, L. N., Greenman, G. D., Schröter, D. C., Hobson, K. A., ... & Vo, A. T. (2016). Does research on evaluation matter? Findings from a survey of American Evaluation Association members and prominent evaluation theorists and scholars. American Journal of Evaluation, 37(2), 159-173.

Coryn, S. (2004). Encyclopedia of evaluation. Sage publications.

Council, C. (2007). Report of the Academic Competitiveness Council.

Council, T. A., & National Academies of Sciences, Engineering, and Medicine. (2016). Science teachers' learning: Enhancing opportunities, creating supportive contexts. National Academies Press.

Creswell, J. W. (2002). A framework for design research: qualitative, quantitative, and mixed methods approaches (pp. 3–23). In: Thousand Oaks: Sage Publications.

Creswell, J. W. (2014). Research design: qualitative, quantitative, and mixed methods approaches (4th ed.). Sage publications.

Creswell, J. W., & Clark, V. L. P. (2017). Designing and conducting mixed methods research. Sage publications.

Creswell, J. W., & Poth, C. N. (2016). Qualitative inquiry and research design: Choosing among five approaches. Sage publications.

Creswell, J. W., Fetters, M. D., & Ivankova, N. V. (2004). Designing a mixed methods study in primary care. *The Annals of Family Medicine*, *2*(1), 7–12.

Dasgupta, N., & Stout, J. G. (2014). Girls and women in science, technology, engineering, and mathematics: STEMing the tide and Broadening participation in STEM careers. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 21–29.

Donaldson, S. I., & Lipsey, M. W. (2006). Roles for theory in evaluation practice. Handbook of Evaluation. Thousand Oaks, Calif.: Sage, forthcoming.

Dugger, W. E. (2010). Evolution of STEM in the United States. 6th biennial international conference on technology education research,

Durbin, D. R., House, S. C., Meagher, E. A., & Rogers, J. G. (2019). The role of mentors in addressing issues of work–life integration in an academic research environment. Journal of Clinical and Translational Science, 3(6), 302-307.

El-Deghaidy, H., & Mansour, N. (2015). Science teachers' perceptions of STEM education: Possibilities and challenges. *International Journal of Learning and Teaching*, *1*(1), 51–54.

Ellis, J., Fosdick, B. K., & Rasmussen, C. (2016). Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PloS one*, *11*(7), e0157447.

Elsea, J. K., & Mason, R. C. (2008). The Use of Federal Troops for Disaster Assistance: Legal Issues. In: Library of Congress

Estrada, M., Eroy-Reveles, A., & Matsui, J. (2018). The influence of affirming kindness and community on Broadening participation in STEM career pathways. *Social issues and policy review*, *12*(1), 258.

Fealing, K. H., Lai, Y., & Myers Jr, S. L. (2015). Pathways vs. pipelines to Broadening participation in the STEM workforce. *Journal of Women and Minorities in Science and Engineering*, *21*(4).

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, *5*(1), 80–92.

Fetterman, D. M., & Wandersman, A. (Eds.). (2005). Empowerment evaluation principles in practice. Guilford Press.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational researcher*, *31*(8), 4–14.

Finelli, S. M. (2012). A Utilization Focused Evaluation: Aligning ISTE Standards Within an Educational Technology Master's Program (Doctoral dissertation, Walden University).

Fitzpatrick, J., Christie, C., & Mark, M. M. (2009). Evaluation in action: Interviews with expert evaluators. Sage.

Flynn, S. (2011). Recalibrating homeland security: Mobilizing American society to prepare for disaster. *Foreign Aff.*, *90*, 130.

Franco, M. S., & Patel, N. H. (2017). Exploring student engagement in STEM education: An examination of STEM schools, STEM programs, and traditional schools. *Research in the Schools*, *24*(1), 10–30.

Frechtling, J. (2010). The 2010 User-Friendly Handbook for Project Evaluation. In. Washington, DC: National Science Foundation (NSF).

Garibay, C., & Teasdale, R. M. (2019). Equity and evaluation in informal STEM education. *New Directions for Evaluation*, *2019*(161), 87–106.

Gee, J. P. (2004). *An introduction to discourse analysis: Theory and method*. Routledge.

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social problems, 12*(4), 436–445.

Gonzalez, H. B. (2012). *An analysis of STEM education funding at the NSF: Trends and policy discussion*. Congressional Research Service Washington, DC.

Gonzalez, H. B., & Kuenzi, J. J. (2012). *Science, technology, engineering, and mathematics (STEM) education: A primer*. Washington, DC: Congressional Research Service, Library of Congress

Granovskiy, B. (2018). Science, Technology, Engineering, and Mathematics (STEM) Education: An Overview. CRS Report R45223, Version 4. Updated. *Congressional Research Service*.

Grant, K. E., Farahmand, F., Meyerson, D. A., Dubois, D. L., Tolan, P. H., Gaylord-Harden, N. K., ... & Duffy, S. (2014). Development of cities mentor project: An intervention to improve academic outcomes for low-income urban youth through instruction in effective coping supported by mentoring relationships and protective settings. Journal of prevention & intervention in the community, 42(3), 221-242.

Greene, J. C. (2007). *Mixed methods in social inquiry* (Vol. 9). John Wiley & Sons.

Greene, J. C., Boyce, A., & Ahn, J. (2011). A values-engaged, educative approach to evaluating educational programs. *A guidebook for practice. Champaign, IL: Department of Educational Psychology, University of Illinois at Urbana-Champaign*.

Greene, J. C., DeStefano, L., Burgon, H., & Hall, J. (2006). An educative, values-engaged approach to evaluating STEM educational programs. *New Directions for Evaluation, 2006*(109), 53–71.

Greenseid, L. O., & Lawrenz, F. (2011). Using citation analysis methods to assess the influence of science, technology, engineering, and mathematics education evaluations. *American Journal of Evaluation*, *32*(3), 392–407.

Groff, J. (2013). Technology-rich innovative learning environments. OCED CERI Innovative Learning Environment project, 2013, 1-30.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, *2*(163–194), 105.

Guba, E. G., & Lincoln, Y. S. (2001). Guidelines and checklist for constructivist (aka the fourth generation) evaluation.

Guillemin, M., & Gillam, L. (2004). Ethics, reflexivity, and "ethically important moments" in research. *Qualitative Inquiry*, *10*(2), 261–280.

Gullickson, A. R., & Hanssen, C. E. (2006). Local Evaluation in Multisite STEM Programs: Relating Evaluation Use and Program Results. *New Directions for Evaluation*, *109*, 87–103.

Hall, J. N., Ahn, J., & Greene, J. C. (2012). Values engagement in evaluation: Ideas, illustrations, and implications. *American Journal of Evaluation*, *33*(2), 195–207.

Hall, J. N., Freeman, M., & Colomer, S. E. (2020). Being Culturally Responsive in a Formative Evaluation of a Professional Development School: Successes and Missed Opportunities of an Educative, Values-Engaged Evaluation. *American Journal of Evaluation*, *41*(3), 384–403.

Hatry, H. P., & Newcomer, K. E. (2004). Pitfalls of evaluation. *Handbook of practical program evaluation*, 547–570.

Havice, W., Havice, P., Waugaman, C., & Walker, K. (2018). Evaluating the Effectiveness of Integrative STEM Education: Teacher and Administrator Professional Development. *Journal of Technology Education*, *29*(2), 73–90.

Heil, D. R., Pearson, G., & Burger, S. E. (2013). Understanding integrated stem education: Report on a national study. 2013 ASEE Annual Conference & Exposition,

Hoeg, D. G., & Bencze, J. L. (2017). Values underpinning STEM education in the USA: An analysis of the Next Generation Science Standards. *Science Education*, *101*(2), 278–301.

Holloman, T., Lee, W. C., London, J., Halkiyo, A., Jew, G., & Watford, B. (2018). A historical and policy perspective on Broadening participation in STEM: Insights from national reports (1974–2016). 2018 CoNECD—The Collaborative Network for Engineering and Computing Diversity Conference,

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. Handbook of practical program evaluation, 281–317.

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. Handbook of practical program evaluation, 281-317.

Hrabowski III, F. A. (2012). Broadening participation in the American STEM workforce. *BioScience*, *62*(4), 325–326.

Huffman, D., & Lawrenz, F. (2006). Editors' notes. *New Directions for Evaluation*, *2006*(109), 1–5. https://doi.org/https://doi.org/10.1002/ev.174

Huffman, D., Lawrenz, F. P., & Lawrenz, F. P. (2006b). Critical issues in STEM education evaluation.

Huffman, D., Lawrenz, F., Thomas, K., & Clarkson, L. (2006). Collaborative evaluation communities in urban schools: A model of evaluation capacity building for STEM education. New Directions for Evaluation, 2006(109), 73-85.

Huffman, D., Lawrenz, F., Thomas, K., & Clarkson, L. (2006a). Collaborative evaluation communities in urban schools: A model of evaluation capacity building for STEM education. *New Directions for Evaluation*, *2006*(109), 73–85.

Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. American Journal of Evaluation, 29(3), 358-368.

Irwanto, I., Saputro, A. D., Widiyanti, W., Ramadhan, M. F., & Lukman, I. R. (2022). Research trends in STEM education from 2011 to 2020: A systematic review of publications in selected journals. International Journal of Interactive Mobile Technologies (iJIM), 16(5), 19-32.

James, S. M., & Singer, S. R. (2016). From the NSF: The National Science Foundation's investments in Broadening participation in science, technology, engineering, and mathematics education through research and capacity building. *CBE—Life Sciences Education*, *15*(3), fe7.

Johnson, J., Hall, J., Greene, J. C., & Ahn, J. (2013). Exploring alternative approaches for presenting evaluation results. *American Journal of Evaluation*, *34*(4), 486–503.

Karl, T. R., Melillo, J. M., Peterson, T. C., & Hassol, S. J. (2009). *Global climate change impacts in the United States*. Cambridge University Press.

Katzenmeyer, C., & Lawrenz, F. (2006). National Science Foundation Perspectives on the Nature of STEM Program Evaluation. *New Directions for Evaluation*, *109*, 7–18. .

Kivunja, C., & Kuyini, A. B. (2017). Understanding and applying research paradigms in educational contexts. International Journal of higher education, 6(5), 26-41.

Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. American Journal of Evaluation, 33(3), 307-338.

Langdon, D., McKittrick, G., Beede, D., Khan, B., & Doms, M. (2011). STEM: Good Jobs Now and for the Future. ESA Issue Brief# 03-11. *US Department of Commerce*.

Lawrenz, F., & Huffman, D. (2004). Using multi-site core evaluation to provide" scientific" evidence [Core Evaluation of the Collaboratives for Excellence in Teacher Preparation Program]. *The Canadian Journal of Program Evaluation*, *19*(2), 17.

Lawrenz, F., & Huffman, D. (2006). Methodological pluralism: The gold standard of STEM evaluation. *New Directions for Evaluation*, *2006*(109), 19–34.

Li, Y., Wang, K., Xiao, Y., Froyd, J. E., & Nite, S. B. (2020). Research and trends in STEM education: a systematic analysis of publicly funded projects. In: SpringerOpen.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. sage.

MacDonald, A., Huser, C., Sikder, S., & Danaia, L. (2020). Effective early childhood STEM education: Findings from the Little Scientists Evaluation. *Early Childhood Education Journal*, *48*(3), 353–363.

MacDonald, J., & Bourke, R. (2017). Reflecting on evaluation practice by considering an educative values-engaged approach: How would it have changed this utilization-focused evaluation? *Eval Matters*, *3*, 130–155.

MacIsaac, D. (2019). Student STEM Content Video Projects for Preservice Teachers: A Drop-in Course Module.

Malyn-Smith, J. (2014, March). A program director's guide to evaluating STEM education programs: Lessons learned from local, state, and national initiatives. In Society for Information Technology & Teacher Education International Conference (pp. 1774-1775). Association for the Advancement of Computing in Education (AACE).

MaxQDA. (2020). MaxQDA [Software]. Version 12. URL: https://www.maxqda.com/products/maxqda-analyze-texts-and-data.

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational researcher*, *33*(2), 3–11.

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

McCloskey, R. (2008). A guide to discourse analysis. *Nurse researcher*, *16*(1).

McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., ... & Barmer, A. (2019). The Condition of Education 2019. NCES 2019-144. National Center for Education Statistics.

Melillo, J. M., Richmond, T., & Yohe, G. (2014). Climate change impacts in the United States. In *Third national climate assessment* (Vol. 52).

Merriam, S. (2009). Qualitative research: A guide to design and implementation San Francisco: John Willey & Sons Inc.

Mertens, D. M., & Hopson, R. K. (2006). Advancing evaluation of STEM efforts through attention to diversity and culture. *New Directions for Evaluation*, *2006*(109), 35–51.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). Qualitative data analysis: A methods sourcebook. 3rd. In: Thousand Oaks, CA: Sage.

Miller, D. I., & Wai, J. (2015). The bachelor's to Ph. D. STEM pipeline no longer leaks more women than men: A 30-year analysis. *Frontiers in Psychology*, *6*, 37.

Mills, S. (2004). *Discourse*. Routledge.

Moon, N. W., Utschig, T. T., Todd, R. L., & Bozzorg, A. (2011). Evaluation of Programmatic Interventions to Improve Postsecondary STEM Education for Students with Disabilities: Findings from SciTrain University. *Journal of Postsecondary Education and Disability*, *24*(4), 331–349.

Morell, J. A. (2010). Evaluation in the face of uncertainty: Anticipating surprise and responding to the inevitable. Guilford Press.

National Academies of Sciences Engineering and Medicine. (2018). Indicators for monitoring undergraduate STEM education. In: National Academies Press.

National Center for Science and Engineering Statistics. (2019). Women, minorities, and persons with disabilities in science and engineering.

National Science Foundation. (2008). Broadening participation at the National Science Foundation: A framework for action. In. Arlington, VA: National Science Foundation (NSF).

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). Planning and designing useful evaluations. Handbook of practical program evaluation, 7-35.

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (Eds.). (2015). Handbook of practical program evaluation. San Francisco, CA: Jossey-Bass & Pfeiffer Imprints, Wiley.

Niess, M., & Gillow-Wiles, H. (2013). Developing asynchronous online courses: Key instructional strategies in a social metacognitive constructivist learning trajectory. *Journal of Distance Education*, *27*(1), 1–23.

Noy, C. (2008). Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. *International journal of social research methodology*, *11*(4), 327–344.

O'Sullivan, T. M., & Ramsay, J. (2015). Defining and distinguishing homeland from national security and climate-related environmental security, in theory, and practice. *Journal of Homeland Security and Emergency Management*, *12*(1), 43–66.

Okrent, A., & Burke, A. (2021, August 31). The STEM Labor Force of Today: Scientists, Engineers, and Skilled Technical Workers. Retrieved from https://ncses.nsf.gov/pubs/nsb20212

O'Leary, Z. (2017). *The essential guide to doing your research project*. Sage.

Pascale, C. M. (2011). Analytic induction. Cartographies of Knowledge: Exploring Qualitative Epistemologies, 39–76.

Patton, M. Q. (2010). Developmental evaluation: Applying complexity concepts to enhance innovation and use. Guilford press.

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.

Petrosino, A. J., Sherard, M. K., & Tharayil, S. A. (2020). The education philosophy, theories, and models that enable STEM policy integration. In *STEM in the Technopolis: The Power of STEM Education in Regional Technology Policy* (pp. 51–63). Springer.

Pouezevara, S., Mekhael, S., & Darcy, N. (2014). Planning and evaluating ICT in education programs using the four dimensions of sustainability: A program evaluation from Egypt. International Journal of Education and development using ICT, 10(2), 120-141.

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. American journal of evaluation, 29(4), 443-459.

Radloff, J., & Guzey, S. (2016). Investigating preservice STEM teacher conceptions of STEM education. *Journal of Science Education and Technology*, *25*(5), 759–774.

Rahman, M. S. (2020). The advantages and disadvantages of using qualitative and quantitative approaches and methods in language “testing and assessment” research: A literature review.

Rauschecker, A. M. (2011). Visual Cortical Circuitry for Building Word Representations. Stanford University.

Reeves, P. M., Bobrownicki, A., Bauer, M., & Graham, M. J. (2020). Communicating complex STEM program evaluation to diverse stakeholders. *CBE—Life Sciences Education*, *19*(2), es4.

Reis, J., Solovey, E. T., Henner, J., Johnson, K., & Hoffmeister, R. (2015). ASL CLeaR: STEM education tools for deaf students. Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility,

Richards, L., & Morse, J. M. (2012). *Readme first for a user’s guide to qualitative methods*. Sage.

Rizzolo, S., DeForest, A. R., DeCino, D. A., Strear, M., & Landram, S. (2016). Graduate student perceptions and experiences of professional development activities. Journal of Career Development, 43(3), 195-210.

Robinson, B. M. (2013). The impact of professional development on teacher effectiveness (Doctoral dissertation, Nova Southeastern University).

Roseland, D., Greenseid, L. O., Volkov, B. B., & Lawrenz, F. (2011). Documenting the impact of multisite evaluations on the science, technology, engineering, and mathematics field. *New Directions for Evaluation*, *2011*(129), 39–48.

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2018). Evaluation: A systematic approach. Sage publications.

Rubin, H. J., & Rubin, I. S. (2005). The responsive interview as an extended conversation. *Qualitative interviewing: The art of hearing data*, 108–128.

Sawyer, R. K. (Ed.). (2005). The Cambridge handbook of the learning sciences. Cambridge University Press.

Saxton, E., Burns, R., Holveck, S., Kelley, S., Prince, D., Rigelman, N., & Skinner, E. A. (2014). A common measurement system for K–12 STEM education: Adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking. *Studies in Educational Evaluation*, *40*, 18–35.

Sayler, B., Apaza, J., & Austin, M. (2005). Using data to make a case for mathematics reform within a K–12 district. *Unpublished working paper presented Sept 15*, 2005.

Schwab, K. (Ed.). (2009). The global competitiveness report 2009–2010. World Economic Forum.

Shade, L., Ludden, T., Dolor, R. J., Halladay, J., Reeves, K., Rees, J., ... & Tapp, H. (2021). Using the Consolidated Framework for Implementation Research (CFIR) to evaluate implementation effectiveness of a facilitated approach to an asthma shared decision-making intervention. Journal of Asthma, 58(4), 554-563.

Shragge, E. (2013). Activism and social change: Lessons for community organizing. University of Toronto Press.

Sindico, F. (2007). Climate change: A security (council) issue. *Carbon & Climate L. Rev.*, 29.

Snyder, T. D., De Brey, C., & Dillow, S. A. (2019). Digest of Education Statistics 2017, NCES 2018-070. National Center for Education Statistics.

Ståhl, T., Zaal, M. P., & Skitka, L. J. (2016). Moralized rationality: Relying on logic and evidence in the formation and evaluation of belief can be seen as a moral issue. PloS one, 11(11), e0166332.

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

Stake, R. E. (2010). Qualitative research: Studying how things work.

Staples, L. (2012). Community organizing for social justice: Grassroots groups for power. Social Work with Groups, 35(3), 287-296.

Steinhauser, K. E., & Barroso, J. (2009). Using qualitative methods to explore key questions in palliative care. *Journal of Palliative Medicine*, *12*(8), 725–730.

STEM, T. B. P. I. (2009). PART I: FOCUS ON NSF STEM Broadening participation PROGRAMS. *Framework for Evaluating Impacts of Broadening Participation Projects*, 9.

Stevahn, L., Berger, D. E., Tucker, S. A., & Rodell, A. (2020). Using the 2018 AEA Evaluator Competencies for effective program evaluation practice. New Directions for Evaluation, 2020(168), 75-97.

Strauss, A., & Corbin, J. (1998). Basics of qualitative research: Techniques and procedures for developing grounded theory (2nd ed.). Sage Publications.

Stufflebeam, D. L., & Coryn, C. L. (2014). *Evaluation theory, models, and applications* (Vol. 50). John Wiley & Sons.

Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (Eds.). (2000). Evaluation models: Viewpoints on educational and human services evaluation (Vol. 49). Springer Science & Business Media.

Swindler, S., Hughes, G., Briggs, C., Yamazaki, K., Ohse, D., Pinero, S., & Sagrestano, L. (2002). American Evaluation Association Annual Conference Evaluation: Evaluation 2002 Final Report.

Tanenbaum, C. (2016). STEM 2026: A vision for innovation in STEM education. *US Department of Education, Washington, DC*.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, *27*(2), 237–246.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC medical research methodology, 8(1), 1–10.

Thomas, V. G., & Parsons, B. A. (2017). Culturally responsive evaluation meets systems-oriented evaluation. *American Journal of Evaluation*, *38*(1), 7–28.

Thompson, B., Levy, L., Lambeth, A., Byrd, D., Alcaidinho, J., Radu, I., & Gandy, M. (2016). Participatory design of STEM education AR experiences for heterogeneous student groups: exploring dimensions of tangibility, simulation, and interaction. 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct),

Trochim, W. M. (2009). Evaluation policy and evaluation practice. *New Directions for Evaluation*, *2009*(123), 13–32.

Tyler-Wood, T. L., Cockerham, D., & Johnson, K. R. (2018). Implementing new technologies in a middle school curriculum: a rural perspective. *Smart Learning Environments*, *5*(1), 1–16.

U.S. Department of Education. (2004). Identifying and implementing educational practices supported by rigorous evidence. In *A user-friendly guide*. Washington, DC: U.S. Department of Education.

van den Hurk, A., Meelissen, M., & van Langen, A. (2019). Interventions in education to prevent STEM pipeline leakage. *International Journal of Science Education*, *41*(2), 150–164.

Van Lange, P. A., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. Organizational Behavior and Human Decision Processes, 120(2), 125-141.

Waite, A. M., & McDonald, K. S. (2019). Exploring challenges and solutions facing STEM careers in the 21st century: A human resource development perspective. Advances in Developing Human Resources, 21(1), 3-15.

Wandersman, A., Alia, K., Cook, B. S., Hsu, L. L., & Ramaswamy, R. (2016). Evidence-based interventions are necessary but not sufficient for achieving outcomes in each setting in a complex world: Empowerment evaluation, getting to outcomes, and demonstrating accountability. American Journal of Evaluation, 37(4), 544-561

White, D. W. (2014). What is STEM education and why is it important? *Florida Association of Teacher Educators Journal*, *1*(14), 1–9.

Whitehurst, G. J. (2009). *An overview of federal STEM education programs. Linking Research and Policy to Improve Science, Technology, Engineering and Math Education: A Seminar for Education Committee Chairs* National Conference of State Legislatures, Huntsville, AL.

Wirt, J., Choy, S., Rooney, P., Provasnik, S., Sen, A., & Tobin, R. (2004). The Condition of Education 2004. NCES 2004-077. *US Department of Education*.

WK Kellogg Foundation. (2004). *WK Kellogg Foundation evaluation handbook*. WK Kellogg Foundation.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users*. Sage Publications.

APPENDIX A: IRB APPROVAL LETTER

IRB #: IRB-FY22-414

Title: Discourses of STEM Education Evaluation: Current and Future Perspectives Creation Date: 2-4-2022 End Date:

Status: Approved

Principal Investigator: Adeyemo Adetogun
Review Board: UNC-Greensboro
IRB Sponsor:

Study History

Submission Type Initial	Review Type Exempt	Decision	Exempt
Key Study Contacts			
Member Aileen Reid	Role Co-Principal Investigator	Contact	amreid3@uncg.edu
Member Adeyemo Adetogun	Role Principal Investigator	Contact	aaadetog@uncg.edu
Member Adeyemo Adetogun	Role Primary Contact	Contact	aaadetog@uncg.edu
Member Ayesha Boyce	Role Investigator	Contact	astillma@uncg.edu

Initial Submission

Getting Started

About Cayuse Human Ethics

Cayuse Human Ethics is an interactive web application. As you answer questions, new sections relevant to the type of research being conducted will appear on the left-hand side. Therefore, not all sections may appear. You do not have to finish the application in one sitting. All information can be saved.

Additional information has been added throughout the form for guidance and clarity. That additional information can be found by clicking the question mark it the top-right corner of each section.

For more information about the IRB submission Process, IRB Tracking, and UNCG IRB Tasks,

please refer to the Cayuse Human Ethics Procedures Manual.

Getting Started

Throughout the submission, you will be required to provide the following:

- Detailed Study Information
- Study-related questionnaires
- Informed Consent Forms
- Study Recruitment Materials

UNCG IRB

- You cannot begin data collection until a formal approval letter from the chair of the IRB has been received.
- Please allow for four weeks for IRB review of your submission. For studies requiring full committee review, the UNCG IRB meets regularly throughout the
- year. If your study is funded, please note that it is the responsibility of the Principal Investigator to link your IRB application to your Cayuse SP record.
- For more information regarding the UNCG IRB, consent form templates, and FAQs, visit our website

APPENDIX B: INFORMED CONSENT FORM

STEM Evaluator Informed Consent

UNIVERSITY OF NORTH CAROLINA GREENSBORO

CONSENT TO ACT AS A HUMAN PARTICIPANT

Project Title: Discourses of STEM Education Evaluation: Current and Future Perspectives

Principal Investigator: Adeyemo Adetogun

Faculty Advisors: Dr. Ayesha Boyce (Dissertation Chair) and Dr. Aileen Reid (Dissertation Co-Chair)

Dear STEM Evaluator:

You are invited to participate in a 45–60-minute telephone or ZOOM interview to discuss your experiences of current and future states of STEM education evaluation. The goal of this research is to understand the Discourses of STEM Education Evaluation in terms of current evaluation approaches and methodologies and to identify the gaps in the approaches and methodologies utilized by evaluation practitioners. Adeyemo Adetogun, a doctoral candidate in the Department of Educational Research Methodology, is conducting this study as part of his dissertation. Dr. Ayesha Boyce and Dr. Aileen Reid serve as his advisors.

What are some general things you should know about research studies?

You are being asked to take part in a research study. Your participation in the study is voluntary. You may choose not to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. There may not be any direct benefit to you for being in the research study. There

also may be risks to being in research studies. If you choose not to be in the study or leave the study before it is done, it will not affect your relationship with the researcher or the University of North Carolina at Greensboro. Details about this study are discussed in this consent form. It is important that you understand this information so that you can make an informed choice about being in this research study.

You will be given a copy of this consent form. If you have any questions about this study at any time, you should ask the researchers named in this consent form. Their contact information is below.

What is the study about?

This is a dissertation study. Your participation is voluntary. The major purpose of this study is to explore the Discourses of STEM education evaluation in terms of existing evaluation approaches and methodologies, as well as to identify gaps in evaluation practitioners' approaches and methodologies in STEM evaluation contexts. Within this purpose, the focus is given to how STEM evaluators characterize issues related to evaluation approaches (e.g., values-engaged, educative; culturally responsive; utilization-focused, responsive, etc.) and the varying methodologies which justify the choice of methods used (e.g., case studies, surveys, focus groups, interviews, etc.).

Why are you asking me?

You are being asked to take part in this study because you have been identified as a STEM program evaluator. You must be 18 or older to participate.

What will you ask me to do if I agree to be in the study?

You will be asked to spend 45 minutes to a (1) hour in a taped interview session with the Principal Investigator, answering a set of research questions and describing your evaluation

experience. The research cannot rule out the remote possibility that you may experience minimal stress based upon your experiences as a STEM evaluator engaging with Discourses of STEM education evaluation. However, every effort will be made to always ensure your safety and welfare during the interview process. You will not be required to perform any physical activity during the interview. The research will not engage in any conduct designed to elicit or create a threatening environment for you, the participant.

Is there any audio/video recording?

The researcher will record digital audio of all individual interview sessions conducted during this research study. These recordings will be transcribed and analyzed to pull out common themes and critical statements. Because your voice will be potentially identifiable by anyone who hears the recording, your confidentiality for things you say on the recording cannot be guaranteed although the researcher will try to limit access to the recording as described below. After the study has been completed and findings documented, the digital recordings of your voice will be destroyed. Information collected during this study is strictly confidential unless disclosure is required by law.

What are the risks to me?

The Institutional Review Board at the University of North Carolina at Greensboro has determined that participation in this study poses minimal risk to participants.

If you have questions, want more information, or have suggestions, please contact Adeyemo Adetogun at (aaadetog@uncg.edu or 919-519-8457), Dr. Ayesha Boyce at (Ayesha.Boyce@asu.edu), or Dr. Aileen Reid at (aileen.reid@uncg.edu)

If you have any concerns about your rights, how you are being treated, concerns or complaints about this project, or benefits or risks associated with being in this study please contact the Office of Research Integrity at UNCG toll-free at (855)-251-2351.

Are there any benefits to society as a result of my taking part in this research?

By participating in this study, you will help inform the research literature on STEM education evaluation and Discourses surrounding evaluation approaches and methodologies

Are there any benefits to me for taking part in this research study?

There aren't any direct or indirect benefits for you to participate in this study.

Will I get paid for being in the study? Will it cost me anything?

There are no costs to you, or payments made for participating in this study.

How will you keep my information confidential?

Names and identifying information will be removed from the transcription record. The recording and transcribed interview will be stored in an electronic folder on UNCG Box. Only the Principal Investigator involved in this study will have access to the information. A master list of all people interviewed, and their associated pseudonyms will be maintained and kept in a separate location from recordings and interview transcriptions.

What if I want to leave the study?

Although we would like to hear from all participants, you have the right to refuse to participate or to withdraw at any time, without penalty. If you do withdraw, it will not affect you in any way. The investigators also have the right to stop your participation at any time. This could be because you have had an unexpected reaction, have failed to follow instructions, or because the entire study has been stopped. Choosing not to participate or withdrawing from the study will in no way affect your relationship with the principal investigator.

What about new information/changes in the study?

If significant new information relating to the study becomes available which may relate to your willingness to continue to participate, this information will be provided to you.

Voluntary Consent by Participant:

By verbally saying yes (or no) below, you are agreeing that you read, and fully understand the contents of this document and are openly willing to take part in this study and all your questions concerning this study have been answered. By signing this form, you are agreeing that you are 18 years of age or older and are agreeing to participate.

To provide your consent to participate in the study, please say "Yes" or select "I consent to participate in this study." If you do not provide your consent, please say "No" or select "I do not consent to participate in this study."

- Yes, I consent to participate in the study.
- No, I do not consent to participate in the study.

APPENDIX C: INTERVIEW PROTOCOL

Discourse of STEM Education Evaluation Study

STEM Education Evaluator Interview Guide (General/Background Questions)

This interview is based on a research project for a dissertation. Your contribution is much valued. The study's main goal is to examine STEM education evaluation Discourses regarding the approaches and methodologies utilized by evaluators within the STEM ecosystem. The study focuses on how STEM evaluators characterize Discourses connected to evaluation approaches (e.g., values-engaged, educative; culturally responsive; utilization-focused, responsive, and so on) and numerous methodologies which justify the methods used (e.g., case studies, surveys, focus groups, interviews, etc.).

This research project uses the term “evaluation approaches” to cover a broad range of practices or ideas about how to conduct STEM education evaluation. However, the study recognizes interview respondents (evaluation scholars/practitioners) may embrace a different terminology or choose not to adopt a particular evaluation approach. Consequently, allowance is made to accommodate the interviewee’s diverse forms of expression that align with the Discourse of this study.

For this study, methodologies (or paradigms, algorithms, etc.) refer to a fundamental research strategy that explains the logic or how the research will be carried out and, among other things, identifies the methods that will be utilized in this research. It differs from methods in that it does not put a priority on specific methods (the means or modes of data collection).

Main Research Questions

The two (2) main research questions (RQs) guiding this study are highlighted below:

RQ 1: What were the Discourse of STEM Education evaluation presentations at the annual American Evaluation Association (AEA) conference from 2014-to 2019?

a. What are the main evaluation approaches and methodologies utilized in these presentations?

b. What are the differences between academic versus non-academic STEM education evaluators (practitioners)?

RQ 2: What are STEM education evaluators' thoughts and impressions about the current and future state of the field?

a. To what extent do the initial findings from the STEM evaluation literature of Discourses resonate with STEM education evaluators about the field's current state? b. How and in what ways could findings influence the field of STEM Education evaluation and its foreseeable future?

Interview Research Questions

The below questions are specifically designed and constructed to help facilitate discussion during the interview process.

1. To begin, will you tell me a little about your background and how you became involved with evaluation in STEM education contexts?

   a. How long have you been practicing evaluation in STEM education contexts?

   b. What kinds of education and training do you have in the field of evaluation as it relates to STEM education programs?

2. As an evaluator of STEM education programs, what sectors and contexts are you working in?

   a. Have you previously presented your work(s) at the AEA conferences?

      I. Can you briefly describe some of the work you presented at AEA?

      II. Can you say more about your work in general?

3. As described in my email, I am interested in investigating the Discourses of STEM education evaluation.

   a. In your own experiences, can you describe your exposure to, or knowledge of evaluation approaches appropriate for evaluations of STEM education programs?

I. What approaches are you most comfortable using, and how do you decide which approaches to apply to a given evaluation context?

II. What evaluation approaches would you prioritize for the STEM education evaluation field and why?

b. In your own experiences, can you describe your exposure to or knowledge of evaluation methodologies appropriate for evaluations of STEM education programs?

I. What methodologies are you most comfortable using, and how do you decide which methodologies to apply to a given evaluation context?

II. What evaluation methodologies (that inform your choice of methods) would you prioritize for the STEM education evaluation field and why?

4. Tell me a little about the direction you see the field of evaluation going as it relates to evaluations of STEM education programs? More specifically in terms of a) evaluation approaches, and b) evaluation methodologies?

5. What are the main issues you focus on addressing in your evaluation practice? (Give examples)

6. What is your evaluation strategy in response to the Covid-19 pandemic, and how has it altered as a result of the ongoing pandemic?

That concludes my interview: Is there anything else you'd like to add or say before we wrap things up?

APPENDIX D: RECRUITMENT EMAIL

Recruitment Email for STEM Evaluators

Hello [insert name],

My name is Adeyemo Adetogun, and I am a doctoral candidate in the educational research methodology department at the University of North Carolina Greensboro (UNCG). I am contacting you to invite you to participate in a study I am conducting as part of my dissertation under the supervision of Dr. Ayesha Boyce and Dr. Aileen Reid.

My dissertation is titled "Discourses of STEM Education Evaluation: Current and Future Perspectives." I am reaching out to you because [insert one of the following reasons]

a) You are a member of the American Evaluation Association (AEA),
b) I have read your evaluation abstract [insert title] that was posted to www.eval.org]

I am writing to invite you to participate in a 45–60-minute telephone or ZOOM meeting interview online to discuss your experiences of current and future states of STEM education evaluation. The study has been approved by the UNCG Institutional Review Board, and all data obtained through this study will be completely confidential. Participants must be 18 years of age or older. Additional details about the study are provided below.

I would greatly appreciate your participation in this study. Please reply to this email or contact me at 919.519.8457 if you are willing to participate.

Thank you for your consideration,

Adeyemo Adetogun
aaadetog@uncg.edu

Additional information about the study:

Study Purpose: The purpose of this study is to explore the Discourses of STEM education evaluation in terms of existing evaluation approaches and methodologies, as well as to identify gaps in evaluation practitioners' approaches and methodologies in STEM evaluation contexts.

Assessment of Risks and Benefits: The Institutional Review Board at the University of North Carolina Greensboro has determined that participation in this study poses minimal risk to participants. The potential benefit of participating in this study is contributing to research knowledge by helping to inform the research literature on STEM education evaluation and Discourses surrounding evaluation approaches and methodologies.

Audio recording: I am requesting to audio record this interview for transcription purposes. Recordings will be transcribed in full, and transcripts will be stored on a secure, password-protected computer for onward analysis. After the study has been completed and findings documented, the digital recordings of your voice will be destroyed. Information collected during this study is strictly confidential unless disclosure is required by law.

Confidentiality: All data will be confidential, and no identifying information will be reported. I will use pseudonyms in place of your name, the name of your organization (if applicable), and the name of the program or intervention you evaluated.

Results: Results from this study will be used for my dissertation and for publications in academic journals.

APPENDIX E: STEM EDUCATION EVALUATOR DEMOGRAPHIC DATA

Table E19. STEM Education Evaluators Interviewed: by Race/Ethnicity, Gender and Sector, N = 13

Race/Ethnicity	Gender	Academic (n = 6)	Private (n = 5)	Public (n = 1)	Private/ Academic (n = 1)	Total (%)
African American	Man	1				1 (8%)
	Woman	3	1		1	5 (38%)
White	Man		1			1 (8%)
	Woman	1	3			4 (31%)
Hispanic/Latino	Man					
	Woman	1				1 (8%)
Asian	Man			1		1 (8%)
	Woman					

Table E20. STEM Evaluators Interviewed: Years in Evaluation Practice (STEM Evaluation Practitioner in Academic, n = 6; Private, n = 5; Academic/Private, n = 1; Public, n = 1)

Years in Evaluation Practice	Academic	Private	Private/ Academic	Public	Total (%)
0 – 5 years	1				1 (8%)
6 – 10 years	2	2			4 (31%)
11 – 15 years	3		1		4 (31%)
16 – 20 years		2			2 (15%)
21 - 25 years		1		1	2 (15%)

APPENDIX F: LOG FRAME FOR DOCUMENT ANALYSIS

Table F21. Log Frame for Document Analysis of AEA STEM TIG Abstracts

#	Author/s	Title	Abstract	Type of Abstract (Paper/Poster//Roundtable etc.)	STEM Evaluation Approaches	STEM Evaluation Methodologies	Dominant Theme	Others
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
nth								

APPENDIX G: EMERGENT CATEGORIES FROM DOCUMENT ANALYSIS

Table G22. Discourse of STEM Education Evaluation at the AEA Conference between 2014–2019

Discussion: Category	Academic	Private	Public
Most Frequent	Lesson Learned	Lesson Learned	Impacts
	Challenges	Challenges	Lesson Learned
	Mentoring	Impacts	Challenges
	Successes	Implementation	Funding
Medium Frequent	Impacts	Funding	Improvement
	Effectiveness	Equity, Diversity, Inclusion (EDI)	Implementation
	Improvement	Successes	Effectiveness
	Implementation	Improvement	Broadening Participation (BP)
Least Frequent	EDI	Effectiveness	Professional Development (PD)
	PD	PD	Successes
	Funding	BP	Mentoring
	BP	Mentoring	EDI