Journal of MultiDisciplinary Evaluation
Volume 18, Issue 42, 2022

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Equity “On the Sideline”: A Mixed Methods Study of New England Evaluation Practice in 2020

Emily F. Gates
Boston College

Eric Williamson
Boston College

Joseph Madres
Boston College

Jori N. Hall
University of Georgia, Athens

Kayla Benitez Alvarez
Boston College

Background: Centering equity in evaluations is increasingly recognized as an important professional responsibility of evaluators. While some theoretical and practical guidance exists, the evaluation field has limited empirical research on equity within evaluation practice.

Purpose: Drawing on select findings from a larger study, this paper explores whether and how evaluators address inequities and advance equity throughout the phased process of evaluation work.

Setting: The study focuses on American Evaluation Association–affiliated evaluators in the New England region of the United States who work in a variety of areas (e.g., health, education).

Intervention: Not applicable**.**

Research Design: Anchored in a conceptual framework for centering equity within evaluation components, the study used a mixed methods design to generate and integrate data from a researcher-developed online questionnaire administered to a census and snowball sample of practicing evaluators (n = 82) and individual, semi-structured interviews (n = 21). Quantitative and qualitative data were analyzed independently by calculating descriptive statistics (i.e., means and standard deviations, frequencies) and thematically analyzing interview data. An integrated analysis using a joint data display was subsequently undertaken.

Data Collection and Analysis: Not applicable

Findings: Eight findings suggest that, despite participating evaluators’ attempts to center equity, equity remains largely “on the sideline” because evaluators seeking to center equity are working against some conventional professional and methodological norms, within contractual and contextual constraints, and with limited professional preparation and workplace supports.

Keywords: *equity; evaluation practice; equitable evaluation; social justice.*

Introduction

The year 2020 marked an unprecedented public reckoning with the extent to which patterns and dynamics of inequity and injustice have long been and continue to be perpetuated throughout global societies. The United States experienced an upsurge in national conversations about systemic racism and white supremacy culture following the murder of George Floyd and other Black Americans by law enforcement officers. Many professional and academic disciplines and institutions responded with statements of condemnation, explicit expression of antiracist values, and related actions or commitments to action. The American Evaluation Association (AEA)'s board of directors, for example, released a statement regarding racism and inequality, calling the evaluation field and evaluators to educate ourselves and work toward justice (Shanker, 2021). This call aligns with AEA's fifth guiding principle, which states that evaluators should "strive to contribute to the common good and advancement of an equitable and just society" (AEA, 2018). Equity, as it is commonly understood within evaluation, means providing fair and just opportunities for people historically or currently underrepresented, marginalized, or least well-served in society and in the specific context of any given evaluation (AEA, 2018; Farrow & Morrison, 2019).

While the AEA's statement regarding racism and inequality brought attention and urgency to issues of inequity, individuals and agencies had been calling for evaluation to center equity throughout the last decade. UNICEF produced a two-part report on equity-focused evaluation with thought leader commentaries and methodological guidance (Bamberger & Segone, 2011). Donaldson and Picciotto (2016) published Evaluation for an Equitable Society, a collection of essays that explore equity in relation to a variety of issues (e.g., ethics, funding, methods) and the broader sociopolitical contexts of democracy and capitalism. A group of leaders in the philanthropic sector launched the Equitable Evaluation Initiative to shift "the evaluation paradigm so that it becomes a tool for and of equity" (Center for Evaluation Innovation, 2017), garnering funding and partnerships from over a dozen national and regional foundations. The Center for Social Policy released a brief entitled "Putting Equity at the Center of Knowledge Development," making the case that

> equity is not a by-product but an essential element—a value—of thoughtfully considered intervention design, learning agendas, and applied data collection and evaluation. This means that inequities and disparities are the problems to be solved, not simply documented. (Farrow & Morrison, 2019, p. 5)

Additionally, several organizations have produced practical guides for equity in applied research (Andrews et al., 2019; Annie E. Casey Foundation, 2014; Hawn Nelson et al., 2020; Wiggins & Sileo, 2020).

Centering equity in evaluation challenges fundamental thinking about the role of evaluation and evaluators' professional responsibilities. Theoretically, conventional evaluation serves society by taking an objective, non-partisan role fulfilled by expert, scientifically trained professionals answering questions posed by funders, policymakers, and program leaders using rigorous designs and methodologies (e.g., Datta, 2000; Rossi et al., 2019). Historically, those responsible for designing, conducting, and reporting on evaluations have not been professionally obliged or prepared to center equity in their work; equity was not a focal consideration, let alone a priority, in the core professional norms outlined in the AEA's pre-2018 guidelines (AEA, 2011), Joint Committee Standards for Educational Evaluation (Yarbrough et al., 2011), or AEA evaluator competencies (AEA, 2018). Furthermore, conventional evaluation theory and practice deny the inherently political, value-laden nature of evaluation—and oppose evaluators who advocate for specific ideological orientations, such as centering equity, in their work. While individual evaluators may examine whether program or policy outcomes and impacts are equitable across all groups of intended beneficiaries if it is a funder requirement, a stated goal of the intervention, or a personally held stance, addressing matters of equity as a professional responsibility and ethical obligation has only recently become an overarching consideration.

Calling out evaluation as a means to advance equity poses a significant challenge and opportunity. Evaluation exists within a marketplace, and evaluations are shaped by client relations. Cronbach and colleagues (1980) pointed this out over 4 decades ago: since evaluations are funded contracts, "the evaluator's aspiration to benefit the larger community has to be reconciled—sometimes painfully—with commitments to a sponsor and to informants, with the evaluator's political convictions, and with his [sic] desire to stay in business" (p. 6). Moreover, the evaluation field does not have a shared normative vision and wrestles with foundational questions about the values and evaluative criteria that should underlie evaluations (Gates, 2018; Schwandt & Gates, 2021). Greene (2014) contends that "one of the greatest challenges of evaluation practice is making justifiable decisions about whose interests will be served and which values will be advanced in an evaluation study" (p. 196).

The AEA's (2018) guiding principles in no way resolve this issue, framing the evaluator's role as recognizing and balancing interests while making efforts to "address the evaluations' potential risks of exacerbating historic disadvantage or inequity" (para 5). Individual evaluators must thus work out—within their own contracts, contexts, and practices—whether and how exactly to advance equity. This raises the question: What does it mean and look like for evaluators to work toward equity within their evaluation practices? Relevant theories and approaches (e.g., culturally responsive evaluation, transformative evaluation) and existing practical guidance (e.g., Andrews et al., 2019) within the field remain prescriptive—pointing to what one should or could do. Little research has examined equity in evaluation practice, leaving practitioners with a very limited empirical knowledge base for such a foundational issue.

This paper presents results from a larger study of what centering equity means and looks like in evaluation practice from evaluators' perspectives. The study intentionally focused on the New England region of the United States to generate data to guide future capacity building within the region and align with the geographical focus of the funding agency. This focal region allowed an examination of the equity-related practices of predominantly white evaluators, whose privilege affords them the choice to speak and act (or to not speak or act) to advance equity for non-dominant groups. We discuss relevant literature, our conceptual framework, and data generation and analysis, and then present eight overarching findings that together support our interpretation that, despite evaluators' attempts to center equity, equity largely remains "on the sideline." We close with discussion of what our findings mean within evaluation literature and directions for future research.

Equity in Evaluation: Theory, Practice, and Research

This section provides a brief examination and synthesis of evaluation literature that informs centering equity within evaluations.

Equity as a principle has roots within several evaluation theories, models, and approaches. Using Mertens and Wilson's (2018) version of the evaluation theory tree, a visual depiction of the types of theorists and the relationships between them, there is evidence that scholars on the "values" and "social justice" branches discuss equity within their work, with equity prominently and distinctly featured within values-engaged, educative evaluation (Greene, 2014; Greene et al., 2006), culturally responsive evaluation (Hood et al., 2015), and deliberative democratic (House & Howe, 1999) and Indigenous evaluation approaches (LaFrance & Nichols, 2009). Equity also connects to ongoing debates about the role of evaluation within structural racism (Caldwell & Bledsoe, 2019) and what this means for whether and how evaluators should act as agents of social change (Hall, 2020). It is, however, beyond the scope of this paper to contrast equity within each of these evaluation models and theorists.

At the time of this study, practical guidance on centering equity was largely ad hoc, produced by multiple agencies for different purposes, without synthesis, and offering different and at times divergent advice. Guidance includes Child Trends' principles and ways to incorporate a racial and ethnic equity perspective throughout stages of the research process (Andrews et al., 2019); the Annie E. Casey Foundation's (2014) guidance for incorporating diversity, equity, and inclusion into organizations; Hawn Nelson and colleagues' (2020) toolkit for centering racial equity throughout data sharing and integration; and the Forum for Youth Investment's brief description of principles for incorporating equity in evaluation policy within federal agencies (Wiggins & Sileo, 2020). While such guidance makes clear that there is no single approach to equity-oriented evaluation, we draw from this guidance to offer an initial conceptualization of equity in evaluation. Centering equity within evaluation practice means putting the equitable, fair, and just treatment of those involved in and affected by an intervention at the core of evaluation design, implementation, and reporting, as well as front and center in negotiations with evaluation funders about criteria used to judge the value of a policy or program. This conceptualization may sound clear, but it does leave considerable room for interpretation and differing applications within practice.

Within research on evaluation, little work has directly focused on the topic of equity in evaluation. Although equity has been discussed as a social, stakeholder, and evaluator value and potential evaluative criterion domain (Teasdale, 2021), there has been little empirical research on values and valuing generally (Coryn et al., 2017), and none on equity specifically. Prior research on evaluation has examined the relationship between theory and practice, identifying a gap between what's espoused and what's done (Christie, 2003; Christie & Lemire, 2019). Studies have also pointed to the growing influence of the evaluation marketplace (Lemire et al., 2018) and to how contextual factors shape valuing in the public interest (Julnes, 2012). Several recent studies with evaluators of color link evaluators' racial identities and lived experiences to their perspectives on social justice within evaluation (Reid et al., 2020). This research reveals challenges evaluators might face when translating theory and guidance into practice within contractual and contextual circumstances. It also raises questions about whether and how white evaluators feel responsibility to advance equity within their practices, and what assumptions they hold about the extent to which they can and should be part of this work.

Conceptual Framework

For this study, we conceptualize evaluation practice according to evaluation process components. To identify these, we first reviewed the Better Evaluation Rainbow Framework (2014), Centers for Disease Control and Prevention Framework (1999), and Culturally Responsive Evaluation (Hood et al., 2015). These frameworks share ways of conceptualizing practice focused on discrete evaluations following a relatively linear sequence. We then iteratively organized descriptions of equity-oriented evaluation from existing guidance and clustered operational practices into components/phases. This led us to a conceptualization comprised of eight practice components (see Figure 1 and, for a brief description of each component, Table 1). This visual is offered as a heuristic to organize data collection and analysis, rather than as a prescriptive guide for designing or conducting equity-oriented evaluation

Figure 1. Conceptual Framework for Equity-Oriented Evaluation

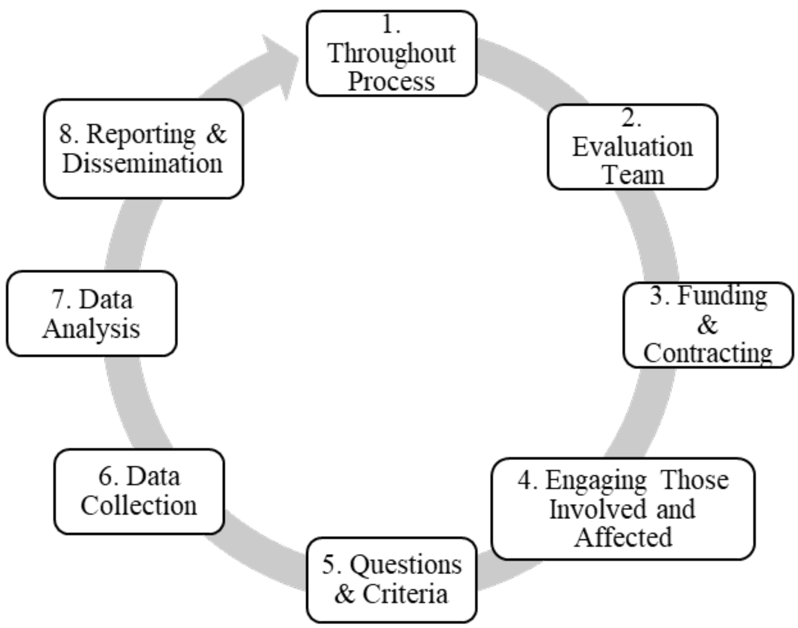

Table 1. Conceptual Framework for Equity-Oriented Evaluation—Component Descriptions

Component	Description
Throughout process	Practices that are carried out within and across components
Evaluation team	Composition of evaluation team leadership and members, and critical self-reflection by evaluator(s) on positionality and assumptions
Funding and contracting	Evaluation team prioritization and negotiation within requests for proposals and requirements from funders/commissioners
Engaging those involved and affected	Whether and how evaluation teams engage groups involved and affected by the intervention, especially in terms of power, culture, and language
Questions and criteria	Whether questions and criteria address equity, and whether intended beneficiaries are involved in shaping them
Data collection	Modification of data collection instruments, especially in terms of culture and language
Data analysis	Kinds of analyses conducted, especially in terms of disaggregation and critical analyses
Reporting and dissemination	Accessibility of evaluation findings and reporting

Design and Methods

Mixed Methods Study Context

This paper presents research conducted within a larger mixed methods study of equity-oriented practices among evaluators in New England. The study addressed five focal areas: (a) characteristics of evaluators (e.g., demographics, experience) and evaluation providers (e.g., firm size, mission); (b) evaluators' understanding/conceptualization/perception of equity within their practices; (c) the extent to which evaluators pursue equity-related practices; (d) supports and barriers to equity-oriented evaluation; and (e) ways to strengthen capacity for equity-oriented evaluation. The study used a mixed methods sequential, explanatory design (Creswell & Plano Clark, 2017) in which findings from individual, semi-structured interviews were used to triangulate and elaborate upon initial results from a questionnaire. Descriptions of the full study and results are publicly available in unpublished reports (Gates et al., 2021) and recent (Gates et al., 2022) and forthcoming manuscripts.

Research Questions

This article reports results from the third research question of the larger study: To what extent and in what ways do evaluators center equity within the design and conduct of evaluations?

Questionnaire Data Generation and Analysis

This section describes questionnaire development, sampling and implementation, and data analysis.

Instrument Development. At the time of the study, there were no instruments available to assess the constructs and items of interest. Therefore, we developed an instrument drawing on existing questionnaires and using a multistep process to establish face and construct validity.

We began by reviewing select literature, guidance, and instruments about culturally responsive and equitable evaluation, and surveys of evaluation practice (AEA, 2011; Andrews et al., 2019; Center for Evaluation Innovation, 2017; Farrow & Morrison, 2019; Hood et al., 2015; Nielsen et al., 2018; Public Policy Associates, Inc., 2015; SPEC Associates, 2018; Westaby et al., 2019). Then we developed a matrix outlining the components of evaluation practice in our conceptual framework, listed existing items for each, and drafted items for missing topics. We revised the item set to provide adequate coverage, use consistent language, and make items comprehensible. We subsequently invited four expert reviewers (Olson, 2010)—senior evaluators specializing in culturally responsive and equitable evaluation—to rate the clarity of and need for each item, and to suggest any missing items. Based on this feedback, we added items to fill gaps, cut redundant items, and clarified language where necessary. We then conducted cognitive interviews (Peterson et al., 2017) with six evaluators engaged in culturally responsive and/or equity-focused evaluation work outside the region, and further revised the instrument.

This process resulted in a three-section questionnaire addressing evaluator backgrounds and characteristics, equity-oriented evaluation practices, and evaluation provider characteristics.[1] This paper focuses on results from the section focused on equity-oriented evaluation practices, and Table 2 provides sample items from this section that speak to specific components of the conceptual framework.

Table 2. Sample Items from each Equity-Oriented Evaluation Framework Component

Component	Sample Item
How often do you / does your team do each of the following?	
Evaluation team	• Have shared lived experiences (e.g., life histories, cultural perspectives, identity-based social positioning) with the population(s) of focus in the evaluation. • Conduct evaluations as part of a racially and/or ethnically diverse team.
Funding and contracting	• Prioritize requests for proposals that have purposes and/or questions focused on inclusion, diversity, and/or equity. • Negotiate with funding agencies to have purposes and/or questions focused on inclusion, diversity, and/or equity.
Engaging those involved and affected	• Work to mitigate power imbalances in the evaluation context. • Challenge any assumptions that the intended beneficiaries lack the ability to achieve because of their culture (i.e., deficit assumptions).
Setting questions and criteria	• Include intended beneficiaries in decision-making about the evaluation purpose, approach, and/or questions. • Include intended beneficiaries in selecting criteria / definitions of success.
Data collection	• Design and/or modify data collection to be appropriate for the culture(s) and language(s) of the people of whom the questions are being asked. • Use systems thinking and/or systems methods (e.g., social network analysis, causal loop diagrams, agent-based modeling, critical systems heuristics).
Data analysis	• Include intended beneficiaries in interpretation of data and sense-making. • When possible, disaggregate data by key demographic differences.
Reporting and dissemination	• Present evaluation results in formats accessible to the intended beneficiaries of the intervention. • Make evaluation results, in some form, publicly accessible.

[1] For a copy of the instrument, please contact the first author.

Sample and Questionnaire Administration. Without a list of evaluators in New England, we opted to focus on evaluators who were AEA members and affiliates, combining a census with snowball sampling (Creswell & Plano Clark, 2017). We obtained a list of AEA members who indicated living in New England (i.e., in Massachusetts, Connecticut, Vermont, Rhode Island, New Hampshire, or Maine), sent the survey invitation to this entire list (n = 258), and invited recipients to forward the invitation to colleagues. Additionally, we sent the invitation out through two national listservs (the Center for Culturally Responsive Evaluation and Assessment and the Annie E. Casey's Foundation's Leaders in Equitable Evaluation and Diversity network) and two regional networks (Greater Boston Evaluation Network and Vermont Evaluation Network). Sixty-seven of 258 evaluators on the AEA list completed the questionnaire, for an initial response rate of 26%. An additional 15 responses came from snowball sampling, for a total of 82 responses. Table 3 compares select demographic characteristics of the survey sample and the individuals on the AEA list to generate a rudimentary understanding of sample representativeness.

Table 3. Questionnaire Sample vs. AEA Population Characteristics

Item	*n*	Sample %	*N*	Population %
Race/Ethnicity	82[2]	100%[1]	258[3]	100%
White	76	93%	164	64%
Asian	4	5%	13	4%
Black / African American	3	4%	14	5%
Hispanic/Latinx	2	2%	7	3%
American Indian, Native American, or Alaskan Native	–	–	2	1%
Multiracial	1	1%	20	8% "Other"
(No response)	–	–	38	15%
Sex	82	100%	258	100%
Female	65	79%	170	66%
Male	14	17%	50	19%
Prefer not to answer / no answer	2	2%	38	15%
Intersex	1	1%	–	–
State	80	100%	258	100%
Massachusetts	35	44%	137	53%
Vermont	16	20%	29	11%
Connecticut	9	11%	29	11%
New Hampshire	9	11%	29	11%
Maine	8	10%	13	5%
Rhode Island	3	4%	21	9%

[2] Indicates a select-all-that-apply question, so % is out of 82 for each response option.
[3] More than one response option was not permitted for the population information.

Sample characteristics roughly align with the AEA population data, with the exception of an oversampling of White and female respondents. Most questionnaire respondents identify as White (93%) and as women (80%) and are between the ages of 30 and 64 (91%). Fewer respondents identify as racial minorities (12%), as men (16%), and as gender non-binary (4%). About one fifth identify as Lesbian, Gay, Bisexual, and/or/ Queer or Questioning (LGBQ+)[4] and 6% as having a disability. Most respondents (95%) hold a doctoral degree, master's degree, or professional degree (e.g., M.S.W., J.D., M.D.). Respondents report an average of 15 years of evaluation experience, with more than half serving as external evaluators (59%), about one third as internal evaluators (35%), and the remaining percentage indicating other roles. Respondents primarily work in Massachusetts (44%), followed by Vermont (20%), Connecticut and New Hampshire (11% each), and Rhode Island and Maine (5% each).

Analysis. We first cleaned any accidental or fraudulent responses in Qualtrics, downloaded all responses into Excel (.xlsx) format, and de-identified the dataset. All Likert-style questions were coded: almost always as 4, often as 3, sometimes as 2, and never or rarely as 1. "Unsure," "outside my scope," and "not familiar with this" response options were coded nominally. We then generated descriptive statistics, including means, standard deviations, and frequencies, for items and component scores, and reported these in tables aligned with the conceptual framework.

Interview Data Generation and Analysis

This section describes interview protocol development, interviewee selection and data generation, and data analysis.

We drafted a protocol to complement and expand on questionnaire results, piloted it with two interviewees engaged in culturally responsive and/or equity-focused evaluation work outside the region, and revised the protocol after both interviews. The final interview protocol consisted of three sections: (a) evaluator background and evaluation work; (b) what in/equity means, whether and how equity is promoted in evaluation practice, and helpful/hindering factors; and (c) reflections on strengthening capacity. Only data regarding equity-related practices are reported here.

We purposefully selected interviewees to maximize variation (Patton, 2015). Among those who indicated on the questionnaire willingness to participate in an interview (33), we utilized several criteria to maximize variation: race/ethnicity, sex/gender, state of residence, LGBQ+ identity, and area of practice. We conducted a total of 21 semi-structured, individual interviews (Patton, 2015) between January 12 and February 12, 2021. Each interview was conducted via Zoom video conferencing for 50 to 60 minutes. Interviewees each received a $40 Amazon gift card. We intentionally shared interview topics with interviewees in advance and encouraged reflection prior to the interviews. One member of the research team led each interview, asking questions and probing for further detail, while another researcher took notes. During interviews, the interviewer periodically restated interviewee responses as a form of member checking (Creswell & Poth, 2018).

Mirroring characteristics of the questionnaire respondents, most interviewees identify as White (81%) and as female (87%). Due to our purposeful sampling, a larger proportion identify as racial/ethnic minorities (19%), as male (19%), and as LGBQ+ (24%). About half are based in Massachusetts (52%), followed by Maine (14%) and New Hampshire (14%). Most are external evaluators (48%), followed by internal (33%) and a mix of both (15%). About half are independent evaluators (52%). Most categorize their responsibilities/role as senior-level (67%). Interviewees most frequently practice in social service contexts (43%), followed by public health and/or health service contexts (38%). See Table 4 for interviewee sample characteristics.

Table 4. Interviewee Sample Characteristics

Item	*n*	%
Race/Ethnicity	21	100%
White	17	81%
Asian	2	9%

[4] The "T" for transgender was excluded because this refers to gender and not sexual orientation or identity.

Black / African American	2	10%
Sex	21	100%
Female	15	71%
Male	4	19%
Prefer not to answer / no answer	1	5%
Intersex	1	5%
State	21	100%
Massachusetts	11	52%
Maine	3	14%
New Hampshire	3	14%
Vermont	2	10%
Connecticut	1	5%
Rhode Island	1	5%
LGBQ+ identification	21	100%
No	16	76%
Yes	5	24%
Evaluator role	21	100%
External	10	48%
Internal	7	33%
A mix of both	3	15%
Area of practice	21[5]	100% [a]
Social services	9	43%
Public health and/or health services	8	38%
Workforce development	6	29%
Pre-K–12 education	6	29%
Higher education	5	24%
International development and aid	5	24%
Informal education (e.g., out-of-school, museums)	4	19%
Organizational development and learning	3	14%
Criminal justice	3	14%
Public policy and analysis	2	10%
Something else [coastal communities, environmental and social development, housing and homelessness, participatory grant making, special education]	5	24%
Position	21	100%
Senior-level evaluator	14	67%
Mid-level evaluator	3	14%
Junior-level evaluator	0	0%
Something else [assistant professor, director of learning and evaluation, independent evaluation consultant]	4	19%
Firm Size	21	100%
Large	6	29%

[5] Indicates a select-all-that-apply question, so % is out of 21 for each response option.

Mid-sized	4	19%
Small / independent evaluator	11	52%

We audio recorded interviews and used an online AI service (https://sonix.ai) to generate transcripts. We reviewed and corrected each transcript for AI mistakes, prepared an annotated transcript, and shared the annotated versions with the interviewees, requesting that they review and make any changes—a second form of member checking (Creswell & Poth, 2018). To analyze the interviews, we used a multistep coding process in the research software program Dedoose (https://www.dedoose.com/). The lead researcher reviewed a random selection of five transcripts and developed an initial codebook composed of deductive codes identified from the research and interview questions, as well as inductive codes (Miles et al., 2014). Two other researchers piloted this codebook to code three interviews each. We met to discuss and revise each code, add additional codes, and finalize the shared codebook (Smagorinsky, 2008). We then checked for intercoder reliability and agreement by having two researchers independently code one quarter of the transcripts, then review the coding for discrepancies. Two researchers then coded all transcripts. Then the lead researcher conducted a thematic analysis, using a constant comparative method for excerpts within and between codes to develop a higher-level categorization of the data (Corbin & Strauss, 2015). We used a shared audit trail (Creswell & Poth, 2018) to track steps in the analysis.

Integrated Analysis

The conceptual framework provided a basis for integrating results within each category. This fulfilled our mixed methods purpose and design to use interviews to complement and expand on questionnaire results. To conduct the integrated analysis (Fetters et al., 2013), we used a combination of a joint data display (Bazeley, 2012) and narrative juxtaposition of results to examine whether and how interview data confirmed, challenged, or expanded questionnaire results. At first, our interview themes did not neatly align with the conceptual framework components, as we had organized relevant inductive codes into beginning, middle, and end phases of an evaluation. Therefore, we reorganized and sometimes reworded interview themes to more clearly align with the language of each evaluation component. We then developed overarching, integrated findings based on questionnaire results (i.e., component means, highest and lowest item scores) and interview themes using a joint display and results narrative (see Table 5).

Findings and Discussion

Here we report eight findings, each grounded in questionnaire and interview evidence, followed by discussion of each finding in the context of prior research and guidance. We use "respondents" to refer to evaluators who completed the questionnaire, "interviewees" to refer to those who participated in interviews, and "participating evaluators" as a term that includes both respondents and interviewees. See Table 5 for the joint display and Table 6 for questionnaire results regarding equity-related practices.

Table 5. Joint Data Display / Results Summary Table

Integrated Findings	Select quantitative results (mean; standard deviation)	Qualitative themes and quotations
1. Throughout evaluation process		
Equity remains largely "on the sideline." Evaluators try to center equity where they can, but find some practices "aspirational" given contractual constraints.	<u>Highest Phases:</u> • Evaluation reporting and dissemination (M = 2.77; SD = 0.77) • Engaging those involved and affected (M = 2.73; SD = 0.61) • Data analysis (M = 2.72; SD = 0.63) <u>Lowest Phases:</u> • Funding and contracting (M = 2.22; SD = 0.56) • Evaluation team (M = 2.48; SD = 0.60)	Throughout phases of an evaluation, evaluators discussed "aspirations" to center equity, with occasional changes to practice given contractual constraints. "I think it's these questions around equity that we have good intentions, but they're always on the sideline. They're always on the back burner because the questions from the funder take so much time." "I would like my values, my goals, my personal theories around the stuff that I do … align with the actual work that I'm doing. But they're not. And I think I might have just given up on having them aligned." "Some of this is aspirational … helping reframe the question and get really clear on how we are limiting ourselves from the start is an important piece of it."
2. Team (M = 2.48; SD = 0.60)		
Evaluators typically do not work on diverse teams. Some ask "Who can I bring?" to expand perspectives and practice critical reflexivity.	<u>Highest Items:</u> • Critically reflect on how I/we relate to others with different cultural backgrounds, identities, and/or perspectives than my/our own (M = 2.98; SD = 0.80) <u>Lowest Items:</u> • Conduct evaluations as part of a racially and/or ethnically diverse team (M = 2.24; SD = 1.00)	Interviewees emphasized wanting to work on teams with colleagues who bring diverse perspectives and lived experience, with some making efforts to hire new evaluators and build partnerships. "I've learned that who is on your evaluation team and what experiences they bring is really, really important. So, thinking about who can I bring that will offer different perspectives and as close as perspectives to the community that the work is to serve as possible."

	• Have shared lived experiences (e.g., life histories, cultural perspectives) with the population(s) of focus in the evaluation (*M* = 2.05; *SD* = 0.74) • Conduct evaluations with leads (PI's, supervisors) who identify as racial and/or ethnic minorities (*M* = 1.83; *SD* = 0.95)	Some interviewees talked about critical reflexivity as a strategy to expand perspectives and mitigate biases, for example, by having team discussions and conducting equity audits of evaluation plans and processes. "We've started an equitable evaluation committee … we do monthly readings and discussions … and project review process.… we talk about how the project is doing equity-wise and how we could do better."
3. Funding and Contracting (*M* = 2.22; *SD* = 0.56)		
Contracts restrict equity, requiring evaluators to "push a lot." Strategies include choosing RFPs carefully, making equity explicit in proposals, having up-front conversations, and budgeting for equity.	<u>Highest Items:</u> • Have to work within specific purposes and questions set by the funder/commissioner (*M* = 2.86; *SD* = 0.87) • Prioritize requests for proposals that have purposes and/or questions focused on inclusion, diversity, and/or equity. (*M* = 2.52; *SD* = 1.02) <u>Lowest Items:</u> • Have to work within a specific design and methods set by the funder/commissioner (*M* = 2.15; *SD* = 0.84) • Work with stakeholders who request culturally responsive and/or equity-focused evaluations (*M* = 1.97; 0.74)	Interviewees shared strategies to expand opportunities to advance equity within funding and contracting: • Choose RFPs carefully to align with values • Make equity explicit in proposals • Have up-front conversations with funders and intervention leaders • Budget for equity and quality "Starting with applying for an RFP, thinking about the evaluation, how early on evaluation is part of the work and even resisting within RFPs things that we know aren't aligned with our values and responding in ways that are intentionally shaped." "In my organization, one of the most central … tenets of what we do is focusing on racial equity. So … people are thinking about it: 'How do we put this in here when we write proposals?' " "Evaluators are not always in a really powerful spot.… So wherever I can, I take what I would term a 'ground-up approach,' where I try to have in-depth conversations to try to understand the meaning behind those goals and impacts." "I've started to budget for more participation at the design phase, so that's just straight up when I write a proposal now.… There's

		money behind that. There's money for me to do at least a few interviews with people who are being affected by the projects."
4. Engaging Those Involved and Affected (*M* = 2.73; *SD* = 0.61)		
Evaluators strive to respect intended beneficiaries but rarely mitigate power imbalances.	Highest Items: • Highlight the strengths of the intended beneficiaries of the evaluation (*M* = 3.09; *SD* = 0.80) • Work to build trust with stakeholders from minoritized or marginalized communities (*M* = 3.01; *SD* = 0.86) • Minimize any potential for the evaluation to exacerbate disadvantage or inequity (*M* = 2.96; *SD* = 0.88) Lowest Items: • Work to mitigate power imbalances in the evaluation context (*M* = 2.45; *SD* = 0.86) • Consult with cultural brokers or translators to mediate between the cultures of stakeholders and the evaluation team (*M* = 2.26; *SD* = 0.93)	Interviewees talked about ways they tried to respect the cultures and perspectives of intended beneficiaries. "I think we try to think about those things like giving people the respect and dignity of being human beings and not looking at them as research subjects or people that can't be trusted to make decisions." Interviewees shared how they struggled to challenge power imbalances between funders, evaluators, leaders, and beneficiaries and that they rarely involved those most affected by an intervention in evaluation decision-making. "We really struggle with the participatory aspect, in terms of, are there members of affected communities that are part of designing and conceptualizing our evaluations? No, not really."
5. Questions and Criteria (*M* = 2.61; *SD* = 0.63)		
Evaluators sometimes work within questions and criteria that address equity. Some evaluators "reframe" or "tweak" pre-set questions and criteria, but typically with little involvement from intended beneficiaries.	Highest Items: • Examine the problem the intervention seeks to address and how that problem is defined (*M* = 3.08; *SD* = 0.90) • Include evaluation question(s) about the outcomes and/or impacts of an intervention on different populations. (*M* = 3.05; *SD* = 0.92) • Include evaluation question(s) related to equity, inclusion, and/or diversity. (*M* = 2.72; *SD* = 0.88)	Most interviewees discussed working within questions set by funders, although some described efforts to "reframe" or "tweak" these questions. "I've been able to choose sometimes the types of questions I have a little more wiggle room on … I try to push the bounds of what they see as reasonable … encouraging them into the right direction and then making them feel like they've come up with a question or idea."

Lowest Items:

- Include intended beneficiaries in selecting criteria / definitions of success (M = 2.29; SD = 0.90)
- Include evaluation question(s) about how an intervention impacts or addresses underlying systemic drivers of inequity (M = 2.29; SD = 1.01)
- Include intended beneficiaries in decision-making about the evaluation purpose, approach, and/or questions (M = 2.28; SD = 0.88)

"How to tweak those evaluation questions so they're not, especially in international development, a rote regurgitation of the OECD DAC criteria. How do you change these evaluation questions so that they are more inclusive of aspects of an intervention at the local level?"

6. Data Collection (M = 2.64; SD = 0.55)

Evaluators try to "avoid top-down" data collection.

Strategies include tailoring instruments through participant consultation and pilots, involving data collectors with shared lived experience, and maximizing benefits and minimizing burdens for participants.

Highest Items:

- Design and/or modify data collection to be appropriate for the culture(s) and language(s) of the people of whom the questions are being asked (M = 3.04; SD = 1.05)

Lowest Items:

- Use systems thinking and/or systems methods (e.g., social network analysis, causal loop diagrams, agent-based modeling, critical systems heuristics) (M = 2.25; SD = 1.09)

Interviewees talked about avoiding "top-down" data collection, for example, by:

- Tailoring instruments through participant consultation and piloting
- Involving data collectors with shared lived experience
- Maximizing benefit and minimizing burden for participants

"Designing the instruments, we do participatory design of the tool, even implementation of it. We try to avoid this big, dramatic top-down thing."

"In data collection, we try as much as we can to think about who is represented on the team and who we're talking to. We don't have a large team. So sometimes we're limited in what we can do there. But it's not necessarily a person who looks like the people we're talking to, but sometimes someone who understands."

"Traditionally, our model has been very much about how we reduce burden on data collection.… we've tried to shift our model so it's less about reducing burden and more about maximizing benefit."

7. Data Analysis (*M* = 2.72; *SD* = 0.63)		
Evaluators try to disaggregate data but also struggle as "tools aren't really set up" to examine inequities.	<u>Highest Items:</u> • When possible, disaggregate data by key demographic differences (*M* = 3.36; *SD* = 0.81) • Look for potential negative consequences of the intervention (*M* = 3.10; *SD* = 0.91) • Look for disparities in access to program services among subgroups of intended beneficiaries (*M* = 3.00; *SD* = 0.98) • Examine differential experiences of participants with/during the intervention (*M* = 2.99; *SD* = 0.85) <u>Lowest Items:</u> • Look for whether and/or how the intervention contributes to systemic and/or structural change on a local, regional, or larger scale (*M* = 2.39; *SD* = 0.91) • Look for whether and/or how the intervention shifts power to intended beneficiaries and their communities (*M* = 2.21; *SD* = 0.96) • Assess whether interventions redistribute resources to those most marginalized and/or disadvantaged (*M* = 2.19; *SD* = 0.97)	Interviewees expressed awareness of the importance of disaggregating data, with some doing so regularly, but most saw methodological challenges to disaggregation and more critical analyses of inequities. "We are always disaggregating. Of course, we pay attention to all the rules about suppressing numbers depending on the '*n*' and that kind of thing … and we always want to understand what it [a demographic variable used to disaggregate] means." "I feel like the analysis phase could be really useful for looking at inequities if the tools are set up to analyze … but I think the tools aren't really set up that way, and then the analysis phase doesn't bode well for looking at inequities." "There's a lot of excuses I find in why we can't look at data with an equity lens because the sample size is too small or … we can't disaggregate because that's too identifying."
8. Reporting and Dissemination (*M* = 2.77; *SD* = 0.77)		
Evaluators disseminate reports in multiple, accessible formats and "ideally" discuss results and co-develop recommendations.	<u>Highest Items:</u> • Present evaluation results in formats accessible to the intended beneficiaries of the intervention (*M* = 3.04; *SD* = 0.97) <u>Lowest Items:</u>	Interviewees provided examples of ways they disseminate evaluation results in multiple formats and across stakeholder groups and, if they didn't, spoke about how they'd like to do so. "We really make efforts to share back data, whether it's in flyers or digested ways of sharing the data back. Sometimes it'll be in a

	• Include intended beneficiaries in interpretation of data and sense-making (M = 2.49; SD = 0.99)	community monthly meeting with the community partners and residents. It's very collaborative." "Ideally, we'd be able to go back to the villages where we collected information. And that's something that generally is not budgeted for.... I've always been disappointed that it's not something that we could do more actively." Some interviewees discussed how they "ideally" would embed processes for shared sense-making, with several providing examples of doing so. "In an ideal world, and sometimes this happens and sometimes it doesn't, is to bring in program participants to look at that and interpret with us." "We had a data party with all of the folks who had contributed to the study along the way. It was optional, but people could join and it was the first draft of the findings and they were into different breakout rooms in Zoom and used a jam board to discuss: What are your key takeaways? What questions do you have and what recommendations do you have?... Those were the recommendations that I put in the final report."

Table 6. Questionnaire Results[6]

Item	Mean (*SD*)	*Never or rarely* (1)	*Sometimes* (2)	*Often* (3)	*Almost always* (4)	Outside scope	Unsure
Evaluation team	2.48 (0.60)						
Critically reflect on how I/we relate to others with different cultural backgrounds, identities, and/or perspectives than my/our own. (*n* = 82)	2.98 (0.80)	4 (5%)	15 (18%)	**42 (51%)**	21 (26%)	0 (0%)	0 (0%)
Identify and challenge my/our personal biases and assumptions. (*n* = 82)	2.93 (0.78)	4 (5%)	16 (20%)	**44 (54%)**	18 (22%)	0 (0%)	0 (0%)
Critically reflect on how we/I share power or alter existing power imbalances between stakeholders. (*n* = 82)	2.72 (0.84)	3 (4%)	**34 (41%)**	27 (33%)	17 (21%)	0 (0%)	1 (1%)
Self-assess my/our privilege and positioning within the evaluation context. (*n* = 82)	2.63 (0.84)	8 (10%)	25 (30%)	**38 (46%)**	11 (13%)	0 (0%)	0 (0%)
Conduct evaluations as part of a racially and/or ethnically diverse team. (*n* = 82)	2.24 (1.00)	22 (27%)	**29 (35%)**	20 (24%)	11 (13%)	0 (0%)	0 (0%)
Have shared lived experiences (e.g., life histories, cultural perspectives) with the population(s) of focus in the evaluation. (*n* = 82)	2.05 (0.74)	16 (20%)	**49 (60%)**	12 (15%)	4 (5%)	0 (0%)	1 (1%)
Conduct evaluations with leads (PIs, supervisors) who identify as racial and/or ethnic minorities. (*n* = 82)	1.83 (0.95)	**37 (45%)**	28 (34%)	9 (11%)	7 (9%)	0 (0%)	1 (1%)
Engaging those involved and affected	2.73 (0.61)						
Highlight the strengths of the intended beneficiaries of the evaluation. (*n* = 82)	3.09 (0.80)	2 (2%)	16 (20%)	**34 (41%)**	27 (33%)	0 (0%)	3 (4%)
Work to build trust with stakeholders from minoritized or marginalized communities. (*n* = 82)	3.01 (0.86)	3 (4%)	20 (24%)	**31 (38%)**	27 (33%)	0 (0%)	1 (1%)
Minimize any potential for the evaluation to exacerbate disadvantage or inequity. (*n* = 81)	2.96 (0.88)	3 (4%)	22 (27%)	**27 (33%)**	25 (31%)	1 (1%)	3 (4%)

[6] Within each section, items are ordered from highest to lowest mean. The mode for each item is shown in bold.

Encourage intervention leaders and/or staff to address root causes of the problem(s) an intervention targets. (*n* = 81)	2.81 (0.90)	3 (4%)	**31 (38%)**	21 (26%)	22 (27%)	2 (2%)	2 (2%)
Conduct evaluation activities in the languages that are relevant/appropriate for the community. (*n* = 81)	2.66 (0.98)	10 (12%)	20 (25%)	**25 (31%)**	16 (20%)	8 (10%)	2 (2%)
Challenge any assumptions that the intended beneficiaries lack the ability to achieve because of their culture (i.e., deficit assumptions). (*n* = 82)	2.65 (0.99)	9 (11%)	**30 (37%)**	20 (24%)	20 (24%)	1 (1%)	2 (2%)
Work to mitigate power imbalances in the evaluation context. (*n* = 82)	2.45 (0.86)	8 (10%)	**38 (46%)**	21 (26%)	11 (13%)	1 (1%)	3 (4%)
Consult with cultural brokers or translators to mediate between the cultures of stakeholders and the evaluation team. (*n* = 82)	2.26 (0.93)	17 (21%)	**33 (40%)**	19 (23%)	9 (11%)	2 (2%)	2 (2%)
Funding and contracting	2.22 (0.56)						
Have to work within specific purposes and questions set by the funder/commissioner. (*n* = 80)	2.86 (0.87)	5 (6%)	18 (23%)	**33 (41%)**	18 (23%)	5 (6%)	1 (1%)
Prioritize requests for proposals that have purposes and/or questions focused on inclusion, diversity, and/or equity. (*n* = 80)	2.52 (1.02)	12 (15%)	**22 (28%)**	19 (24%)	14 (18%)	8 (10%)	5 (6%)
Have an adequate budget that allows time and resources to involve stakeholders in evaluations. (*n* = 80)	2.39 (0.90)	11 (14%)	**33 (41%)**	20 (25%)	10 (13%)	4 (5%)	2 (3%)
Have to work within a specific design and methods set by the funder/commissioner. (*n* = 80)	2.15 (0.84)	17 (21%)	**33 (41%)**	20 (25%)	4 (5%)	5 (6%)	1 (1%)
Negotiate with funding agencies to have purposes and/or questions focused on inclusion, diversity, and/or equity. (*n* = 80)	2.09 (1.00)	**22 (28%)**	**22 (28%)**	14 (18%)	7 (9%)	9 (11%)	6 (8%)
Work with stakeholders who request culturally responsive and/or equity-focused evaluations. (*n* = 79)	1.97 (0.74)	18 (23%)	**42 (53%)**	10 (13%)	3 (4%)	3 (4%)	3 (4%)
Questions and criteria	2.61 (0.63)						
Examine the problem the intervention seeks to address and how that problem is defined. (*n* = 81)	3.08 (0.90)	4 (5%)	17 (21%)	28 (35%)	**31 (38%)**	0 (0%)	1 (1%)
Include evaluation question(s) about the outcomes and/or impacts of an intervention on different populations. (*n* = 80)	3.05 (0.92)	4 (5%)	19 (24%)	25 (31%)	**31 (39%)**	1 (1%)	0 (0%)

Include evaluation question(s) related to equity, inclusion, and/or diversity. (*n* = 81)	2.72 (0.88)	5 (6%)	**31 (38%)**	27 (33%)	18 (22%)	0 (0%)	0 (0%)
Negotiate criteria / definitions of success across different stakeholder groups. (*n* = 81)	2.48 (0.86)	11 (14%)	27 (33%)	**33 (41%)**	8 (10%)	0 (0%)	2 (2%)
Include intended beneficiaries in selecting criteria / definitions of success. (*n* = 81)	2.29 (0.90)	15 (19%)	**33 (41%)**	22 (27%)	8 (10%)	2 (2%)	1 (1%)
Include evaluation question(s) about how an intervention impacts or addresses underlying systemic drivers of inequity. (*n* = 81)	2.29 (1.01)	20 (25%)	**26 (32%)**	21 (26%)	11 (14%)	2 (2%)	1 (1%)
Include intended beneficiaries in decision-making about the evaluation purpose, approach, and/or questions. (*n* = 81)	2.28 (0.88)	16 (20%)	**31 (38%)**	26 (32%)	6 (7%)	1 (1%)	1 (1%)
Data collection	2.64 (0.55)						
Design and/or modify data collection to be appropriate for the culture(s) and language(s) of the people of whom the questions are being asked. (*n* = 80)	3.04 (1.05)	7 (9%)	16 (20%)	15 (19%)	**33 (41%)**	5 (6%)	4 (5%)
Use systems thinking and/or systems methods (e.g., social network analysis, causal loop diagrams, agent-based modeling, critical systems heuristics). (*n* = 79)	2.25 (1.09)	22 (28%)	**27 (34%)**	11 (14%)	15 (19%)	2 (3%)	2 (3%)
Data analysis	2.72 (0.63)						
When possible, disaggregate data by key demographic differences. (*n* = 79)	3.36 (0.81)	1 (1%)	13 (16%)	21 (27%)	**43 (54%)**	1 (1%)	0 (0%)
Look for potential negative consequences of the *intervention*. (*n* = 79)	3.10 (0.91)	2 (3%)	22 (28%)	19 (24%)	**34 (43%)**	1 (1%)	1 (1%)
Look for disparities in access to program services among subgroups of intended beneficiaries. (*n* = 80)	3.00 (0.98)	5 (6%)	21 (26%)	19 (24%)	**31 (39%)**	3 (4%)	1 (1%)
Examine differential experiences of participants with/during the intervention. (*n* = 80)	2.99 (0.85)	2 (3%)	22 (28%)	**29 (36%)**	25 (31%)	0 (0%)	2 (3%)
Analyze how interactions between race, ethnicity, class, gender, sexual orientation, etc. influence differential outcomes. (*n* = 80)	2.84 (0.84)	2 (3%)	**27 (34%)**	**27 (34%)**	19 (24%)	4 (5%)	1 (1%)
Look for potential negative consequences of the *evaluation*. (*n* = 79)	2.49 (1.14)	19 (24%)	**21 (27%)**	16 (20%)	20 (25%)	1 (1%)	2 (3%)

Examine the extent to which intended beneficiaries were actively involved in the planning and implementation of program activities. (*n* = 80)	2.41 (0.89)	12 (15%)	**28 (35%)**	27 (34%)	8 (10%)	4 (5%)	1 (1%)
Look for whether and/or how the intervention contributes to systemic and/or structural change on a local, regional, or larger scale (*n* = 80)	2.39 (0.91)	11 (14%)	**35 (44%)**	19 (24%)	11 (14%)	4 (5%)	0 (0%)
Look for whether and/or how the intervention shifts power to intended beneficiaries and their communities. (*n* = 80)	2.21 (0.96)	17 (21%)	**33 (41%)**	12 (15%)	10 (13%)	7 (9%)	1 (1%)
Assess whether interventions redistribute resources to those most marginalized and/or disadvantaged. (*n* = 80)	2.19 (0.97)	19 (24%)	**26 (33%)**	16 (20%)	8 (10%)	8 (10%)	3 (4%)
Reporting and dissemination	2.77 (0.77)						
Present evaluation results in formats accessible to the intended beneficiaries of the intervention. (*n* = 80)	3.04 (0.97)	5 (6%)	21 (26%)	20 (25%)	**34 (43%)**	0 (0%)	0 (0%)
Make evaluation results, in some form, publicly accessible. (*n* = 79)	2.75 (0.93)	6 (8%)	**27 (34%)**	24 (30%)	20 (25%)	2 (3%)	0 (0%)
Include intended beneficiaries in interpretation of data and sense-making. (*n* = 79)	2.49 (0.99)	13 (16%)	**29 (37%)**	22 (28%)	15 (19%)	0 (0%)	0 (0%)

Throughout Evaluation Process

Together, questionnaire and interview results characterize evaluators experiencing equity "on the sideline," with attempts by some to center equity to varying degrees in each evaluation component. On average, respondents engage in equity-related activity in the evaluation process between sometimes (a score of 2) and often (a score of 3). Most often, respondents center equity during the following components of the evaluation: evaluation reporting and dissemination (M = 2.77), engaging those involved and affected by the intervention (M = 2.73), and data analysis (M = 2.72). Evaluators reported being least likely and/or able to center equity in evaluation funding and contracting (M = 2.22) and within composition of an evaluation team (M = 2.48). Qualitative data provide examples of evaluators' limited capacity to prioritize equity within contracts, as the interviewee quoted in this paper's title reveals:

> I think it's [through] these questions around equity that we have good intentions, but they're always on the sideline. They're always on the back burner because the questions from the funder take so much time.

While interviewees would like their work to advance equity, they reflected on how their practices fell short of their ideals: "Some of this is aspirational … helping reframe the question and get really clear on how we are limiting ourselves from the start is an important piece of it." Another interviewee despairingly reflected, "I would like my values, my goals, my personal theories around the stuff that I do … [to] align with the actual work that I'm doing. But they're not. And I think I might have just given up on having them aligned." This sentiment was echoed, to varying extents, among interviewees who shared disconnects between how they would like to center equity throughout the evaluations they are part of and their actual day-to-day practice tendencies.

Such disconnects between aspirations and practice reflect research findings regarding the theory–practice relationship in evaluation: evaluators do not directly apply evaluation models/approaches to practice but do so adaptively and strategically within specific circumstances (Christie, 2003; Christie & Lemire, 2019). This indicates the value of a mix of high-level frameworks and principles (e.g., Dean-Coffey, 2017; EEI & GEO, 2021), reflective questions (e.g., Cerna et al., 2021), decision trees, and other forms of guidance, rather than prescriptive models. This finding also speaks to the need for broader changes within the evaluation ecosystem (EEI & GEO, 2021; Gates et al., 2022) to influence enabling conditions for equity-oriented work, particularly funding/commissioning.

Evaluation Teams

Who comprises an evaluation team and whether and how this team addresses in/equity is something interviewees considered important. Interviewees spoke about the value of diverse perspectives:

> I've learned that who is on your evaluation team and what experiences they bring is really, really important. So, thinking about who can I bring that will offer different perspectives and as close as perspectives to the community that the work is to serve as possible.

Although interviewees value diverse evaluation teams, respondents reported least often conducting evaluations as part of racially and/or ethnically diverse teams (M = 2.48), being on evaluation teams led by evaluators who identify as racial and/or ethnic minorities (M = 1.83), and working with team members who have shared lived experiences with the populations of focus (M = 2.05). Interviewees further described having limited power to change the composition of teams they are part of, partly due to their seniority and role. Despite their limited power, evaluators described trying to address this issue in two main ways: by asking, "Who can I bring" onto the team to expand perspectives, and by cultivating critical reflexivity. To expand perspectives, interviewees indicated that they augment their teams by creating partnerships to subcontract or collaborate with evaluators outside their own organizations, building in ways to involve local community members as co-evaluators, or, in some cases, offering training and capacity building within an evaluation. To cultivate critical reflexivity, interviewees indicated that they developed practices within their teams and organizations to openly reflect on and critique biases and limitations and identify opportunities for more equitable work. Some even formalized these practices in their organizations. For example, one interviewee described an "equitable evaluation committee" that talks about "how the project is doing equity-wise and how we could do better."

These results align with prior research and guidance underscoring the importance of evaluation team composition, critical reflection, and learning. These results also mirror guidance on scrutinizing how privilege shapes evaluators' perspectives (Hall, 2020; Symonette et al., 2020) and intentionally cultivating team cultures that minimize cultural stereotyping and misunderstanding-based biases (AEA, 2011; Hood et al., 2015). Relevant guidance calls for racially and ethnically diverse teams that incorporate multiple perspectives (Andrews et al., 2019; Cerna et al., 2021) with attention to cultural competencies (AEA, 2011) such as language and translation ability (Cerna et al., 2021). Guidance also suggests that "community evaluation committees" (Chicago Beyond, 2018) be established so evaluation teams can receive input grounded in committee members' lived experiences (Coleman et al., 2017), and/or involving community members as co-evaluators. Some research has found that while including community members as evaluators raises capacity-building concerns, benefits arise from just-in-time training (Coleman et al., 2017).

Funding and Contracting

Of all the evaluation components, evaluators reported least often centering equity within funding and contracting (M = 2.22). Respondents must often work within purposes and questions set by the funder/commissioner (M = 2.86) and only sometimes get the opportunity to work with stakeholders who request culturally responsive and/or equity-focused evaluations (M = 1.97). This means that evaluators, in general, experience limited professional autonomy to change the scope of contracts, though some seem to have comparatively more influence over which contracts they as individuals, teams, or organizations work on. Although evaluators center equity least often in this evaluation component, the qualitative data highlight four strategies some interviewees use to advance equity within funding and contracting.

First, some evaluators choose requests for proposals (RFPs) carefully and prioritize those with opportunities to advance equity. For example, one interviewee described her thought process when selecting RFPs: "Starting with applying for an RFP, thinking about the evaluation, how early on evaluation is part of the work and even resisting within RFPs things that we know aren't aligned with our values and responding in ways that are intentionally shaped." Questionnaire results confirm evaluators' privileging of RFPs that advance equity, indicating that respondents frequently prioritize RFPs that have purposes and/or questions focused on inclusion, diversity, and/or equity (M = 2.52).

Second, some interviewees explicitly incorporate equity into proposals as a value of the organization and focus of the design: "In my organization, one of the most central … tenets of what we do is focusing on racial equity. So … people are thinking about it: 'How do we put this in here when we write proposals?' "

Third, some interviewees emphasized early conversations and relationship building with funders/commissioners as crucial:

> Evaluators are not always in a really powerful spot…. So, wherever I can, I take what I would term a "ground-up approach" where I try to have in-depth conversations to try to understand the meaning behind those goals and impacts [prioritized by funders].

Fourth, among interviewees who work on evaluation budgets, several spoke about the costs of equity-oriented evaluation, particularly with respect to participant involvement and qualitative methods, noting that the availability of adequate resources can afford or deny possibilities for equity:

> I've started to budget for more participation at the design phase, so that's just straight up when I write a proposal now … there's money behind that. There's money for me to do at least a few interviews with people who are being affected by the projects.

These findings reflect and expand on prior considerations. While guidance generally suggests critically reviewing language in grant proposals (Public Policy Associates, Inc., 2015), our findings point to specific strategies to formalize these reviews within evaluation teams and organizations, as well as the need for shared language to incorporate equity into proposals. Prior work and our interviews highlight how funders (and evaluation leads) may need to change "what they regard as the most valuable knowledge" (Farrow & Morrison, 2019) and craft budgets accordingly, shifting from the norm of most resources going to data collection, analysis, and reporting to budgeting for participant and community benefits (Chicago Beyond, 2018; EEI & GEO, 2021). The existing literature points to relevant strategies such as budgeting "for greater inclusion" (Lo & Espiritu, 2021, p. 9), incorporating "intangible costs to participants, community organization, and community" (Chicago Beyond, 2018, p. 24), and supporting capacity building (Lo & Espiritu, 2021). To support such shifts, some thought leaders in the field have called for relationship building and dialogue between funders and evaluators (EEI & GEO, 2021). Results indicate the need for nuanced discussion and strategies that consider how interpersonal and organizational power dynamics, seniority, and privilege influence funder–evaluator relationships.

Engaging Those Involved and Affected

Questionnaire respondents and interviewees often work to respect and build relationships with groups intended to benefit from the intervention being evaluated. Respondents report most frequently highlighting the strengths of the intended beneficiaries of the evaluation (M = 3.09), working to build trust with stakeholders from minoritized or marginalized communities (M = 3.01), and minimizing any potential for the evaluation to exacerbate disadvantage or inequity (M = 2.96). Interviewees talked about respecting intended beneficiaries throughout the evaluation process. In relation to data collection, one interviewee talked about seeing themself as a "passive translator … really trying to engage and respect the individual I'm working with." This meant "giving people the respect and dignity of being human beings and not looking at them as research subjects or people that can't be trusted to make decisions." Another interviewee, an internal evaluator, discussed how their staff reframed accessibility to consider the influences of racism on participants' experiences: the "psychological and emotional weight that it places on" participants. This, in turn, broadened how they designed for accessibility of programmatic activities to incorporate physical, cognitive, and emotional considerations.

Respondents reported least frequently working to mitigate power imbalances in the evaluation context (M = 2.45) and consulting with cultural brokers or translators to mediate between the cultures of stakeholders and the evaluation team (M = 2.26). Interviewees shared how they struggled to challenge power imbalances between funders, evaluators, leaders, and beneficiaries. For example, one mentioned the use of "minority" within a federally funded health project: "Why do you keep saying 'minority' if they're not a minority in that community?" This interviewee recalled asking the funder to change the language, being told they couldn't change it, and subsequently wondering if they were "perpetuating [the funder's] approach and language and vision of thinking about things." Most interviewees shared challenges associated with involving those most affected by an intervention in evaluation decision-making: "We really struggle with the participatory aspect, in terms of, are there members of affected communities that are part of designing and conceptualizing our evaluations? No, not really."

These findings reveal an opportunity for further guidance on how evaluators can "balance the interests of the client, other stakeholders, and the common good" (AEA, 2018), mitigate power imbalances (Andrews et al., 2019; Symonette et al., 2020), and prioritize intended beneficiaries and their communities (Hawn Nelson et al., 2020). For evaluators in our study, the ideals of balance and equity can be incompatible, such as when funder and community interests are at odds with evaluators as intermediaries. Furthermore, our results indicate a greater practice of engagement over empowerment, as distinguished by the Annie E. Casey Foundation (2014):

> There is a difference between stakeholder engagement and empowerment. Engagement may simply involve getting input or limited participation. Empowerment involves taking leadership, making decisions and designing solutions and strategies at every phase of social-change efforts. (p. 6)

Setting Questions and Criteria

On the questionnaire, we asked respondents to indicate how often they set or work within various types of evaluation questions and criteria drawing from those emphasized within equity-oriented guidance. Recall from the Funding and Contracting section that respondents often work within purposes and questions set by funders/commissioners. Here they report most often investigating the underlying problem an intervention seeks to address (M = 3.08) and the outcomes and/or impacts of an intervention on different populations (M = 3.05). Somewhat less often they work within questions explicitly related to equity, inclusion, and/or diversity (M = 2.72), and occasionally or rarely they answer questions about how an intervention addresses systemic drivers of inequity (M = 2.29). Similarly, more than half of respondents (58%, or 46 respondents) report using evaluative criteria that "examine equity of the intervention's opportunities, experiences, benefits, and/or results." Considerably more respondents use criteria that address the "effectiveness of the intervention at achieving desired results, outcomes, or objectives" (73; 92%), followed by "relevance of the intervention to the needs, culture, interests, or circumstances of the intended beneficiaries" (65; 82%). Together, these results indicate that respondents sometimes work within questions and criteria that address equity and, more often, address effectiveness and outcomes that may or may not incorporate considerations of inequities. See Table 7 for results on evaluation criteria used; those in italics relate to equity.

Table 7. Questionnaire Results: Evaluation Criteria[7]

Item	Response	*N* (%)
Evaluation criteria	**Effectiveness of the intervention at achieving desired results, outcomes, or objectives**	**73 (92%)**
	Relevance of the intervention to the needs, culture, interests, or circumstances of the intended beneficiaries	65 (82%)
	Intended consequences/effects of the intervention	64 (81%)
	Experience of intended beneficiaries during the intervention	62 (78%)
	Design of the intervention in accordance with relevant theoretical principles, best practices, standards, and/or laws	60 (76%)
	Sustainability of the intervention to continue beyond the start-up period or for the benefits to occur over a longer time frame	56 (71%)
	Alignment of the intervention with larger initiatives, related interventions, funder aims, and/or interconnected problems	53 (67%)
	Unintended consequences / side effects of the intervention	50 (63%)
	Equity of the intervention's opportunities, experiences, benefits, and/or results, with particular consideration to prioritizing marginalized populations	46 (58%)
	Replicability of the intervention to be duplicated or adapted to another context	35 (44%)
	Costs and resource use of the intervention	31 (39%)
	Something else (please specify)	5 (6%)
	We do not use criteria in our evaluations.	0 (0%)
	Total	**79 (100%)**
Other criteria specified in open response	Alignment between principles of the program/strategy and actual practices	1 (20%)
	Attribution (or not) of impact to intervention	1 (20%)
	Satisfaction of participants in the evaluation process/outcomes	1 (20%)
	Social & ecological considerations	1 (20%)
	Utilization focus	1 (20%)
	Total	**5 (100%)**

Many interviewees shared a sense of frustration about having to work within questions, pre-set by funders/commissioners, that exclude or ignore issues of in/equity. Some provided examples of efforts to "reframe" or "tweak" questions, as evident here:

> I've been able to choose sometimes the types of questions I have a little more wiggle room on … I try to push the bounds of what they see as reasonable … encouraging them into the right direction.

Working in international development, another evaluator raised questions and criteria as a frustrating challenge they had little control to address, first commenting on the need to "tweak those evaluation questions" so that they're "not a rote regurgitation of the OECD DAC criteria." They went on to say that the "questions just don't make any sense" and even "the whole program didn't make any sense" from perspectives of local people. "But it's too late at the evaluation stage." Repeating a phrase used in relation to other evaluation components, an interviewee spoke about reframing questions as "aspirational"—a way to create "open-ended" inquiry that could "actually produce nuanced data and nuanced analysis," although they

[7] The mode for each item is shown in bold. Items that pertain to equity are italicized.

added that this is rarely what funders/commissioners seem to want. Other interviewees, on the other hand, did routinely find opportunities to ask "questions that are really exciting, learning questions, to the people who are involved in the work."

Results point to a pattern of little to no involvement from intended beneficiaries in crafting questions and criteria or making decisions about the broader evaluation approach. Respondents reported least often including intended beneficiaries in decision-making about the evaluation approach, purpose, and/or questions (M = 2.28) and in the process of defining evaluative criteria (M = 2.29). One interviewee explained that, in their work, this is because the "scope of work is so diffuse," involving "dozens of sub-grantees and organizations who in turn work with many different communities," with organizational representatives often becoming a "stand-in" for a particular community. Several interviewees spoke about using needs assessments as the bases for identifying questions and criteria. For example, one reflected on their "advocacy role" as an evaluator, taking this to mean talking about "what that population needs." Another emphasized the importance of having a "good balance of people that understand the needs of that particular population" in early conversations about evaluation scope. As these examples illustrate, consideration of intended beneficiaries within questions and criteria tends to be through representatives and not direct involvement.

This finding mirrors suggested research and guidance on equity-oriented questions and criteria and reveals a gap between such guidance and these evaluators' practices. It is evident that equity-oriented work should "illuminate the potential impact of race and racism on the programs that we evaluate and the environments that we engage" (Thomas et al., 2018, p. 514), examine structural inequities (Andrews et al., 2019), and critique normative assumptions underlying programs (Schwandt & Gates, 2016). However, evaluative criteria tend to focus on distribution of resources, opportunities, and outcomes (Giacomini & Hurley, 2008; Teasdale, 2021). While evaluators in our study modify "effectiveness" and "impact" foci to examine disparities, they rarely examine deeper, structural questions about root causes. Additionally, though guidance underscores involvement of intended beneficiaries in shaping an evaluation's scope (Annie E. Casey Foundation, 2014; Andrews et al., 2019; EEI & GEO, 2021; Farrow & Morrison, 2019), we found that evaluators more often tried to anticipate and incorporate what they thought would be beneficial, with infrequent direct inclusion.

Data Collection

Data collection is an area where evaluators push against some conventional ways of working and feel they often have the power and authority to do so. Respondents often design and/or modify data collection to culturally and/or linguistically suit the participants (M = 3.04). They less often use systems thinking and/or systems methods (e.g., social network analysis, causal loop diagrams) in their work (M = 2.25), practices we asked about because systems change (e.g., root causes, structural dynamics) is emphasized within equity-oriented guidance (Petty & Leach, 2020).

This section of the questionnaire contained only two items, revealing the limited emphasis on this component within the existing literature and the guidance resources on which the instrument was based. Yet, interviewees had a lot to say about this component, sharing numerous strategies for avoiding "top-down" data collection, three of which were common among those we spoke with. First, interviewees critiqued conventional modes of administering a research-developed instrument with little contextual alignment and participant input. They emphasized the importance of consulting with participants to ascertain construct validity and relevancy, and the importance of piloting instruments. One interviewee noted, "Whether it's designing the instruments, like designing the focus group protocols, we … do participatory design of the tool, even implementation of it. We try to avoid this big, dramatic top-down thing." Another interviewee focused on language translation: "We do definitely look at instrument development and may need to translate the survey items. In the past, we have definitely translated survey instruments into Spanish." Second, some interviewees perceived a mismatch between the identities and lived experiences of those on their teams and the cultural and linguistic practices of program participants. They emphasized the need to involve data collectors with shared lived experiences, although this depended on team composition, evaluation budget, and other factors.

> In data collection, we try as much as we can to think about who is represented on the team and who we're talking to. We don't have a large team. So sometimes we're limited in what we can do there. But it's not necessarily a person who looks like the people we're talking to, but sometimes someone who understands.

A third set of practices involve reducing the burden and maximizing benefit for participants, such as through fair and meaningful incentives, among other considerations, as this interviewee noted: “We’ve tried to shift our model so it’s less about reducing burden and more about maximizing benefit.” Examples of how evaluators do this include considering accessibility of participation (e.g., timing, transportation) and meaningfulness of incentives (e.g., cash versus food).

These findings reflect emphases within prior guidance on instrument development and shifting from transactional to relational approaches (EEI & GEO, 2021). Cultural competence and responsiveness emphasize “data collection tools that are respectful and responsive to the needs of different groups” (Carter & Hossain, 2020). This includes asking participants to codevelop or vet instruments, translating the language and concepts used, and cognitively testing instruments with the population of interest (Andrews et al., 2019). In theory, it also means addressing “cultural differences in how knowledge is constructed and communicated" (AEA, 2011, p. 8), though this was a lesser emphasis among participating evaluators. Guidance also encourages shifting from transactional to relational approaches to “consider the needs, opportunities and learning that can be of service to nonprofit partners and their constituents, as well as the funder” (EEI & GEO, 2021, p. 28), which connects with interviewees’ emphasis on maximizing benefits for participants while leaving room for further consideration regarding what this shift entails.

Data Analysis

Mirroring existing guidance, our questionnaire contained the most items regarding the data analysis component. Respondents reported often disaggregating data by demographic differences (M = 3.36) but less often (M = 2.84) analyzing how interactions between race, ethnicity, class, gender, sexual orientation, etc. influence differential outcomes. Ways they disaggregate data include looking for disparities in access to services among subgroups of intended beneficiaries (M = 3.00) and examining differential experiences of participants with/during an intervention (M = 2.84). Interviewees expressed awareness of the importance of disaggregating data, and some reported doing so regularly, but most saw methodological challenges:

> We are always disaggregating. Of course, we pay attention to all the rules about suppressing numbers depending on the “n” and that kind of thing … and we always want to understand what it [a demographic variable used to disaggregate] means.

Challenges discussed include limited demographic information, lack of knowledge about how to examine differences, particularly for small sample sizes and from an intersectional lens, and concerns about threats to anonymity. Several interviewees saw these limitations as surmountable excuses and sought strategies to meaningfully examine inequities. This points to a generally felt sentiment that familiar analytic techniques were insufficient:

> I feel like the analysis phase could be really useful for looking at inequities, if the tools are set up to analyze … but I think the tools aren’t really set up that way, and then the analysis phase doesn’t bode well for looking at inequities.

Beyond disaggregation, questionnaire and interview data point to limited use of more critical and in-depth analyses of inequities. Respondents reported often looking for potential negative consequences of an intervention (M = 3.10) and only sometimes looking for potential negative consequences of the evaluation (M = 2.49). Another practice that occurs only sometimes is examining whether and/or how the intervention contributes to systemic and/or structural change on a local, regional, or larger scale (M = 2.39). Least often of all, respondents look at shifts in power to intended beneficiaries and communities (M = 2.21) and redistribution of resources to those most marginalized and/or disadvantaged (M = 2.19). Several interviewees did raise possibilities for more in-depth analyses that examine root causes of inequities and leverage points for change but were unsure about methodological options. One interviewee remarked, “It’s pretty new, in this process of triangulating our data sources, to really figure out what the root causes are or what our intervention points are. It’s been interesting, but it’s difficult.”

These findings indicate the need to train practicing evaluators in recent guidance and developing guidance for structural and systemic analyses. Disaggregation and intersectional analyses are widely valued within equity work (Andrews et al., 2019; Annie E. Casey Foundation, 2014; Hawn Nelson et al., 2020). Yet, much of the how-to guidance is recent (e.g., We All Count’s work for equity in data science), meaning that evaluators in the field need pathways to learn about and practice these techniques. Even so, equity calls for “addressing multiple structural, institutional, interpersonal, and individual causes of inequity (both historic and current)” (Change Elemental, n.d., 1st para.) and identifying forward-looking strategies to adapt to and proactively influence changes to systems (Carden, 2013). In other words, “inequity is rooted in a complex range of political, social, and economic factors” that analyses should identify and interrogate (Hopson et al., 2012)—something evaluators acknowledged but rarely practiced, citing lack of knowledge and skill.

Report Dissemination

Respondents often present reports in formats accessible to intended beneficiaries (M = 3.04) and less often make evaluation results, in some form, publicly accessible (M = 2.75). Many interviewees provided examples of ways they disseminate evaluation results; one noted:

> We really make efforts to share back data, whether it’s in flyers or digested ways of sharing the data back. Sometimes it’ll be in a community monthly meeting with the community partners and residents.... It’s very collaborative.

Another said: “We do a lot of silly stuff and hashtags and infographics and one-pagers and just trying to get folks excited; it really depends on what it is and how actionable it is for people.”

Some respondents include intended beneficiaries in the interpretation and sense-making of data (M = 2.49). Similarly, some interviewees discussed how they embed processes for shared sense-making, involving participants in facilitated processes to interpret and discuss the data ranging from a form of member checking for accuracy—described as asking, “Does it sound right?”—to more involved co-interpretation such as a “data party” and the collaborative generation of takeaways and recommendations.

This last finding raises a question about the differences between what’s good evaluation practice, as understood within professional standards and guidance (AEA, 2018), and what an equity orientation adds. Prior guidance encourages the use of tailored formats, language, and style for different audiences (AEA, 2011) and evaluator facilitation of learning and use (Andrews et al., 2019; Cerna et al., 2021)—approaches clearly commonplace among participating evaluators. However, ensuring that the evaluation makes meaningful differences in the lives of those most vulnerable or identifies pathways to do so while minimizing harm (AEA, 2018) seems to go beyond evaluators’ scopes of work and normative conceptions of professional responsibility.

Practice Implications, Limitations, and Future Research Directions

Practice Implications

Our results underscore the need to bridge current guidance with capacity building for evaluators and changes to the ecosystems that influence how and by whom evaluations are conducted. While evaluators may be aware of the shifts needed in their practice, they generally lack the professional autonomy and methodological and interpersonal skill sets to center equity throughout evaluation processes. Those funding and commissioning evaluations wield considerable power and shape scopes of work, and evaluators are in the business of securing and maintaining funding for their work. This implies that creating the conditions for equity-oriented evaluation will require changes by funding agencies (EEI & GEO, 2021), as well as changes to hiring practices within evaluation organizations (Lo & Espiritu, 2021) and training for practicing evaluators.

Limitations

Limitations of our work include our conceptualization of evaluation, sample characteristics, our questionnaire instrument, and our lack of benchmarks. First, conceptualizing “practice” as evaluations comprised of discrete phases sets boundaries around what equity means and how it can be addressed. While this conceptualization of evaluation facilitated a focus on equity within the design and conduct of evaluations, we encourage work on equity within the evaluation ecosystem, as well (Gates et al., 2022). Second, due to the lack of a representative sample, inferences cannot be generalized to evaluators within the New England region. Moreover, given sample size and characteristics, our quantitative analyses were limited to descriptive statistics. That said, our study provides insights into patterns that may be transferable to other regional evaluation contexts, evaluation workplaces, and evaluators. Third, given the lack of instruments on this topic at the time of study, we developed a new instrument and relied on anchoring the items in the literature and expert reviews for face and construct validity; no psychometric testing was done on the instrument. Fourth, we were limited in our capacity to draw normative interpretations of how well participating evaluators center equity given the lack of shared guidance and benchmarks. Others may render more or less critical interpretations than ours and identify additional implications.

Future Directions for Research on Equity in Evaluation

Several areas need future research. First, research should examine the relationship between factors intrinsic to an evaluation (e.g., evaluators' identities, training, years of experience, areas of practice) and extrinsic factors (e.g., the funding agency, size of the evaluation firm, evaluand's goals), and how these interact to influence the way the evaluation process addresses equity. This would enable the field to develop a richer and more contextually grounded conception of equity-oriented evaluation practice which, in turn, could lead to better guidance tailored to evaluators and the circumstances in which they work. Second, research should construct theories of change to depict the sets of assumptions about which and how specific equity-oriented practices within an evaluation contribute to equity for a particular group, community, or context. This will open possibilities to empirically test and refine these theories and identify the most promising practices. Third, given the necessity of unlearning and learning anew for many evaluators and for altering some professional norms, a line of work should explore what kinds of transformations are unfolding and how best to purposefully cultivate these.

Conclusion

Despite calls and some practical guidance for centering equity over the past decade, the evaluation field has only begun to acknowledge equity as a professional responsibility of evaluators. In this paper, we sought to ground debates in the perspectives of evaluation practitioners to explore how equity gets interpreted and addressed within day-to-day evaluation work. Research on evaluation has grown, but there has been little empirical investigation of values generally (Coryn et al., 2017) or of values related to equity and racial and social justice, more specifically (Reid et al., 2020). Our work moves forward a line of inquiry and sheds some light on equity within evaluation practices in New England during 2020 and early 2021.

Funding

This study was supported by a grant from the Barr Foundation (Grant Number: 20–08544) under the leadership of Yvonne Belanger, Director of Evaluation and Learning.

Acknowledgments

We thank those who participated in pilot testing our questionnaire and interview protocols, those who participated in the study by completing the questionnaire and/or interview, reviewers for their feedback, and Tawnya Fay for her editorial work.

References

American Evaluation Association (AEA). (2011). *American Evaluation Association statement on cultural competence in evaluation*. www.eval.org/About/Competencies-Standards/Cutural-Competence-Statement

American Evaluation Association (AEA). (2018). *American Evaluation Association: Guiding principles for evaluators*. www.eval.org/About/Guiding-Principles

Andrews, K., Parekh, J., & Peckoo, S. (2019). *How to embed a racial and ethnic equity perspective in research: Practical guidance for the research process*. Child Trends.

Annie E. Casey Foundation. (2014). *Race equity and inclusion action guide. Embracing racial equity: 7 steps to advance and embed race equity and inclusion within your organization*. www.aecf.org/resources/race-equity-and-inclusion-action-guide

Bamberger, M., & Segone, M. (2011). *How to design and manage equity-focused evaluations*. UNICEF.

Bazeley, P. (2012). Integrative analysis strategies for mixed data sources. *American Behavioral Scientist*, *56*(6), 814–828. https://doi.org/10.1177/0002764211426330

Better Evaluation Rainbow Framework. (2014). http://betterevaluation.org

Caldwell, L. D., & Bledsoe, K. L. (2019). Can social justice live in a house of structural racism?

A question for the field of evaluation. *American Journal of Evaluation*, *40*(1), 6–18. https://doi.org/10.1177/1098214018815772

Carden, F. (2013). Asking questions in the solution space: Methodological issues in evaluating equity. *Evaluation and Program Planning*, *36*(1), 213–217. https://doi.org/10.1016/j.evalprogplan.2012.03.010

Carter, C. E., & Hossain, F. (2020). *How to apply a diversity, equity, & inclusion lens to strengthen evidence building in the social sector*. Project Evident. https://toolkit.projectevident.org/p/how-to-apply-a-diversity-equity-inclusion-lens-to-strengthen-evidence-building-in-the-social-sector

Center for Evaluation Innovation. (2017). *Equitable evaluation project framing paper*. https://www.equitableeval.org/_files/ugd/21786c_04631f8cfdd44cb4a10a4097bf85acd5.pdf

Centers for Disease Control and Prevention. (1999). Framework for program evaluation in public health. *Morbidity and Mortality Weekly Report*, *48*(RR-11).

Cerna, O., Condliffe, B., & Wilson, A. (2021). Guiding questions for supporting culturally responsive evaluation practices and an equity-based perspective. www.mdrc.org/publication/guiding-questions-supporting-culturally-responsive-evaluation-practices-and-equity-based

Change Elemental. (n.d.). Deep equity. https://changeelemental.org/offerings/deep-equity/

Chicago Beyond. (2018). *Why am I always being researched?* https://chicagobeyond.org/researchequity/

Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation, 97*, 7–36. https://doi.org/10.1002/ev.72

Christie, C. A., & Lemire, S. T. (2019). Why evaluation theory should be used to inform evaluation policy. *American Journal of Evaluation, 40*(4), 490–508. https://doi.org/10.1177/1098214018824045

Coleman, S., Byrd, K., Scaccia, J., Stout, S., Schall, M., Callender, S., Anderson, J., Behrman, N., Budnik, A., Smith, D., Brown, L., Douglas, W., Bussey, R., McDermott, E., Munene, E., Mullin, F., Hatchett, L., Pohorelsky, J., VanLanen, T., Pairolero, & B., Mann, Z. (2017). Engaging community members with lived experience. *SCALE 1.0 Synthesis Reports*. Institute for Healthcare Improvement.

Corbin, J. M., & Strauss, A. L. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). SAGE.

Coryn, C. L. S., Wilson, L. N., Westine, C. D., Hobson, K. A., Ozeki, S., Fiekowsky, E. L., Greenman, G. D., & Shroeter, D. C. (2017). A decade of research on evaluation: A systematic review of research on evaluation published between 2005 and 2014. *American Journal of Evaluation, 38*(3), 329–347.

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. SAGE.

Cronbach, L. J., & Associates. (1980). *Toward reform of program evaluation*. Jossey-Bass.

Datta, L.-E. (2000). Seriously seeking fairness: Strategies for crafting non-partisan evaluations in a partisan world. *American Journal of Evaluation*, *21*(1), 1–14. https://doi.org/10.1177/109821400002100101

Dean-Coffey, J. (2017). Equitable Evaluation Framework™. Equitable Evaluation Initiative. www.equitableeval.org/framework

Donaldson, S. I., & Picciotto, R. (2016). *Evaluation for an equitable society*. Information Age Publishing.

Equitable Evaluation Initiative (EEI), & Grantmakers for Effective Organizations (GEO). (2021). *Shifting the evaluation paradigm: The Equitable Evaluation Framework*. https://www.equitableeval.org/_files/ugd/21786c_7db318fe43c342c09003046139c48724.pdf

Farrow, F., & Morrison, S. (2019). *Placing equity concerns at the center of knowledge development*. Center for the Study of Social Policy. https://cssp.org/wp-content/uploads/2019/05/Putting-Equity-at-the-Center-of-Knowledge-Development.pdf

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—principles and practices. *Health Services Research*, *48*(6, Pt. 2), 2134–2156. https://doi.org/10.1111/1475-6773.12117

Gates, E. F. (2018). Toward valuing with critical systems heuristics. *American Journal of Evaluation*. *39*(2), 201–220. https://doi.org/10.1177/1098214017703703

Gates, E. F., Williamson, E., Madres, J., Benitez Alvarez, K., & Hall, J.N. (2021). *Strengthening capacity and equity in New England evaluation (SCENE) Full Report*. Boston College. https://www.researchgate.net/publication/353804145_Strengthening_Capacity_and_Equity_in_New_England_Evaluation_SCENE_Full_Report

Gates, E. F., Madres, J., Hall, J. N., & Alvarez, K. B. (2022). It takes an ecosystem: Socioecological factors influencing equity-oriented evaluation in New England, U.S., 2021. *Evaluation and Program Planning, 92*, 102068. https://doi.org/10.1016/j.evalprogplan.2022.102068

Giacomini, M., & Hurley, J. (2008). Issues in evaluating equity. In *Health promotion evaluation practices in the Americas* (pp. 285–298). Springer.

Greene, J. (2014). Values-engaged evaluation. In M. Bamberger & M. Segone (Eds.), *Evaluation for equitable development results* (pp. 196–206). UNICEF.

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. https://doi.org/10.1002/ev.178

Hall, J. N. (2020). The other side of inequality: Using standpoint theories to examine the privilege of the evaluation profession and individual evaluators. *American Journal of Evaluation, 41*(1), 20–33. https://doi.org/10.1177/1098214019828485

Hall, M. E. (2020). Blest be the tie that binds. *New Directions for Evaluation, 166*, 13–22.

Hawn Nelson, A., Jenkins, D., Zanti, S., Katz, M., & Berkowitz, E. (2020). *A toolkit for centering racial equity throughout data integration*. Actionable Intelligence for Social Policy. University of Pennsylvania. www.aisp.upenn.edu/centering-equity/

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. In K. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (pp. 281–318). Jossey-Bass.

Hopson, R.K., Kirkhart, K.E., & Bledsoe, K. (2011). Decolonizing Evaluation in a Developing World: Implications and Cautions for Equity-Focused Evaluation. In M. Segone (Ed.), *Evaluation for equitable development results* (pp. 59-82).

House, E., & Howe, K. R. (1999). *Values in evaluation and social research*. SAGE.

Julnes, G. (2012). Developing policies to support valuing in the public interest. *New Directions for Evaluation, 133*, 109–129. https://doi.org/10.1002/ev.20012

LaFrance, J., & Nichols, R. (2009). Indigenous evaluation framework: Telling our story in our place and time. American Indian Higher Education Consortium.

Lemire, S., Nielsen, S. B., & Christie, C. A. (2018). Toward understanding the evaluation market and its industry—advancing a research agenda. *New Directions for Evaluation, 160*, 145–163. https://doi.org/10.1002/ev.20339

Lo, F., & Espiritu, R. (2021). Evaluation is so White: Systemic wrongs reinforced by common practices and how to start righting them. Funder & Evaluator Affinity Network.

Mertens, D. M., & Wilson, A. T. (2018). *Program evaluation theory and practice, second edition: A comprehensive guide*. Guilford Publications.

Miles, M. B., Huberman, A. M, & Saldaña, J. (2014). *Qualitative data analysis: a methods sourcebook* (3rd ed.). SAGE.

Nielsen, S. B., Lemire, S., & Christie, C. A. (2018). The evaluation marketplace and its industry. *New Directions for Evaluation, 160*, 13–28.

Olson, K. (2010). An examination of questionnaire evaluation by expert reviewers. *Field Methods, 22*(4), 295–318. https://doi.org/10.1177/1525822X10379795

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE.

Peterson, C. H., Peterson, N. A., & Powell, K. G. (2017). Cognitive interviewing for item development: Validity evidence based on content and response processes. *Measurement and Evaluation in Counseling and Development, 50*(4), 217–

223. https://dx.doi.org/10.1080/07481756.2017.1339564

Petty, S., & Leach, M. (2020). Systems change & deep equity: Pathways toward sustainable impact, beyond "Eureka!," unawareness & unwitting harm. Change Elemental. https://changeelemental.org/resources/systems-change-and-deep-equity-monograph/

Public Policy Associates, Inc. (2015). *Considerations for conducting evaluation using a culturally responsive and racial equity lens*. http://publicpolicy.com/wp-content/uploads/2017/04/PPA-Culturally-Responsive-Lens.pdf

Reid, A. M., Boyce, A. S., Adetogun, A., Moller, J. R., & Avent, C. (2020). If not us, then who? Evaluators of color and social change. *New Directions for Evaluation, 166*, 23–36. https://doi.org/10.1002/ev.20407

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach* (8th ed.). SAGE.

Schwandt, T. A., & Gates, E. F. (2016). What can evaluation do? An agenda for evaluation in service of an equitable society. In Donaldson, S. and Picciotto, R. (Eds.), *Evaluation for an Equitable Society* (pp. 67-81). Charlotte, NC: Information Age Publishing, Inc.

Schwandt, T.A. & Gates, E.F. (2021). *Evaluating and valuing in social research*. Guilford Press.

Shanker, V. (2021). Statement from the AEA Board of Directors regarding racism and inequality in our society. American Evaluation Association. www.eval.org/Full-Article/statement-from-the-aea-board-of-directors-regarding-racism-and-inequality-in-our-society

Smagorinsky, P. (2008). The method section as conceptual epicenter in constructing social science research reports. *Written Communication, 25*(3), 389–411.

SPEC Associates. (2018). *Survey of small sellers of evaluation services: Print version* [Unpublished survey instrument].

Symonette, H., Miller, R. L., & Barela, E. (2020). Power, privilege, and competence: Using the 2018 AEA Evaluator Competencies to shape socially just evaluation practice. *New Directions for Evaluation, 168*, 117–132.

Teasdale, R. M. (2021). Evaluative criteria: An integrated model of domains and sources. *American Journal of Evaluation*, *42*(3), 354–376. https://doi.org/10.1177/1098214020955226

Thomas, V. G., Madison, A., Rockcliffe, F., DeLaine, K., & Lowe, S. M. (2018). Racism, social programming, and evaluation: Where do we go from here? *American Journal of Evaluation*, 39(4), 514–526. https://doi.org/10.1177/1098214018772910

Westaby, K. A., Williams, T. M., Robinson, N. N., & Connors, E. (2019). *Being responsive: The first assessment of culturally responsive evaluation in Wisconsin: Findings from the 2017 survey*. ¡Milwaukee Evaluation!, Inc.

Wiggins, M. E., & Sileo, A. (2020). What's the role of equity in evaluation policy? The Forum for Youth Investment.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users*. SAGE.