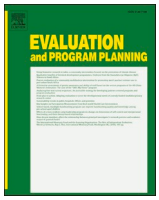
Evaluation and Program Planning 92 (2022) 102068

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan



It takes an ecosystem: Socioecological factors influencing equity-oriented evaluation in New England, U.S., 2021

Emily F. Gates [a,*], Joseph Madres [b], Jori N. Hall [c], Kayla Benitez Alvarez [b]

[a] Measurement, Evaluation, Statistics, and Assessment Department, Lynch School of Education and Human Development, Boston College, USA
[b] Lynch School of Education and Human Development, Boston College, USA
[c] Department of Lifelong Education, Administration, and Policy, University of Georgia, Athens, USA

ARTICLE INFO

Keywords:
Equity
Evaluation practice
Equity-oriented/equitable evaluation
Evaluation ecosystem

ABSTRACT

Evaluation must transform to center equity. Yet, while recent scholarship critiques evaluation at the macro level for reproducing societal inequities and calls the profession and individual evaluators to change, this research overlooks *evaluation ecosystems* – though dynamic interactions among evaluation teams, workplaces, community stakeholders, funders, and informal professional networks form crucial connections between the macro and micro levels and can be spaces for promoting equity within and through evaluations. Addressing this gap, this exploratory study proposes and uses an adapted socioecological framework to organize thematic analysis of data from interviews with evaluators in New England (n = 21) about factors that help and hinder equity-oriented evaluation practices. We identify nine domains and twenty-three factors across macro, meso, and micro levels that influence these evaluators' capacity to practice equity-oriented evaluation in regional, national, and international contexts. The study contributes a framework that future research can adapt to explore the relevance of identified domains and factors to other geographical settings. We also provide questions to guide evaluators, program leaders, and others in reflecting on leverage points for change within their own contexts and outline future directions for research on equity and evaluation.

1. Introduction

Recent scholarship reframes our collective understandings of evaluation concerning historical, systemic, and institutional racism and other forms of oppression (Hall, 2020; Bamberger & Segone, 2011; Caldwell & Bledsoe, 2019; Dean-Coffey, 2018; Donaldson & Picciotto, 2016; Hall, 2018, 2020). Several scholars suggest that the field must transform to center equity through evaluation practice (Bamberger & Segone, 2011; American Evaluation Association AEA, 2018; Donaldson & Picciotto, 2016; Farrow & Morrison, 2019). Some scholars critique the evaluation field's complicity in macro-level societal inequities and injustices (Mathison, 2016; Shanker, 2021) and call for change from the profession (Hall, 2018, 2020) and from individual evaluators and evaluations (Caldwell & Bledsoe, 2019). In modifying its fifth guiding principle, the AEA (2018) argues evaluators should "strive to contribute to the common good and advancement of an equitable and just society." Numerous published guides for centering equity in evaluation provide direction and norms for shifts in evaluation practice (Andrews, Parekh, & Peckoo, 2019; Bamberger & Segone, 2011; Cerna, Condliffe, & Wilson, 2021; Change Elemental, 2021; Farrow & Morrison, 2019; Hawn Nelson, Jenkins, Zanti, Katz, & Berkowitz, 2020; Public Policy Associates Inc., 2015).

Even as equity considerations become established as a professional norm (AEA, 2018), there remain significant challenges to advancing equity through and within evaluations (Giacomini & Hurley, 2008). Evaluators, program leaders, stakeholders, and others must first understand how inequity is systemically produced within the contexts surrounding evaluations, and what it means for evaluations to advance equity within these contexts. Evaluation as a professional practice operates within a marketplace (Lemire, Nielsen, & Christie, 2018a), and evaluators work on teams within organizations that compete for funder contracts. These contracts, the organizations that win the bids (Peck, 2018), and the evaluation teams that carry out the evaluation work all shape how and by whom evaluations get done. The Equitable Evaluation Initiative (EEI & GEO, 2021), in its shared report with Grantmakers for Effective Organizations (GEO)(2021) on EEI's Equitable Evaluation Framework™ (Dean-Coffey, 2017) states that "the parts (of evaluation) must become an integrated ecosystem to allow for and create new and

* Corresponding author.
E-mail addresses: gatesea@bc.edu (E.F. Gates), madres@bc.edu (J. Madres), jorihall@uga.edu (J.N. Hall), benitezk@bc.edu (K.B. Alvarez).

https://doi.org/10.1016/j.evalprogplan.2022.102068
Received 29 September 2021; Received in revised form 4 March 2022; Accepted 5 March 2022
Available online 10 March 2022

emerging ways of evaluative practice that supports individuals, teams and organizations, and is reflected in behaviors and systems" (p. 9). Examining evaluation as an ecosystem offers tremendous potential to expand how we understand the societal, organizational, interpersonal, and intrapersonal factors that influence and shape the ways and extent to which evaluations can center equity. Yet, while exploring evaluation through a socioecological lens can identify the changes needed to best support equity-oriented evaluation, this crucial work is yet to be done.

In this paper, we propose a framework for conceptualizing evaluation as an ecosystem and use it to organize factors identified by evaluators in the New England region of the United States as influencing equity within evaluations they conducted regionally, nationally, and/or internationally. We developed this framework, adapted from Bronfenbrenner's (1981, 1994) ecological systems theory, and identified factors through iterative thematic analysis of data from evaluator (n = 21) interviews conducted in 2021 as part of a mixed methods study. We found evidence supporting the thesis that centering equity in evaluation requires interrelated changes throughout evaluation ecosystems. Our results, organized into socioecological levels and domains, provide a preliminary set of factors that, according to evaluators, influence the potential for evaluations and evaluators to advance equity.

2. Defining equity in evaluation

Evaluation pursuing equity has been described as "equity-focused" (Bamberger & Segone,2011), "centering equity in knowledge development" (Farrow & Morrison, 2019), "equitable" (Dean-Coffey, 2017; EEI & GEO, 2021), "culturally responsive and equitable" (Mendez & Taniuchi, 2020; Anderson & Mastri, 2021), and "culturally responsive and equity-based" (Cerna et al., 2021). In this article, we use "equity-oriented evaluation" to refer to a broad commitment to address inequities and advance various forms of equity through evaluation processes and products. Our study design was informed by the AEA's (2018) definition of equity as "the condition of fair and just opportunities for all people to participate and thrive in society regardless of individual or group identity or difference. Striving to achieve equity includes mitigating historic disadvantage and existing structural inequalities." The analysis section describes how we asked interviewees to define equity within their practice and inductively drew on that information when analyzing their subsequent responses.

3. Reviewing research on equity and evaluation

Much of the academic scholarship on equity and evaluation has focused on theoretical and methodological issues, with little empirical research on equity-oriented evaluation practices. Theoretical writings about equity in evaluation highlight how societal structures and ideologies – neoliberal capitalism (Mathison, 2016), meritocracy (Stake, 2016), settler colonialism (Hopson, Kirkhart, & Bledsoe, 2012), a white racial frame (House, 2017), and systemic racism (Caldwell & Bledsoe, 2019) – influence what/who gets evaluated, by whom, based on whose values, using which method(ologie)s, and with what consequences. Examples of scholars' calls for the evaluation field and profession to shift include incorporating a section on race and racism in the *American Journal of Evaluation* (Hall, 2018) and rethinking philanthropic funding (Center for Evaluation Innovation, 2017; Dean-Coffey, 2018). Evaluation thought leaders have examined critical methodological issues regarding how to center equity (Carden, 2013; Rogers, 2016) and how different evaluation approaches and models can be used to advance equity within individual evaluations (House & Howe, 1999; Greene, 2014, 2016).

While equity has been discussed as a potential evaluative criterion domain (Giacomini & Hurley, 2008; Teasdale, 2021), there has been little empirical research on equity within evaluation practice. Studies present equity within evaluation in different international contexts (Forestieri, 2020) and practice areas (Boyce, 2017; Garibay & Teasdale, 2019), and in relation to Indigenous communities (Letiecq & Bailey, 2004; LaFrance & Nichols, 2009) and dis/ability and access (Lucas, van Wee, & Maat, 2016). Several recent studies with evaluators of color examine culturally responsive evaluation practices within regional evaluation groups (Westaby, Williams, Robinson, & Connors, 2019) and links between the identities of evaluators of color and their personal commitments to social justice within evaluation (e.g., Reid, Boyce, Adetogun, Moller, & Avent, 2020).

Building on this prior work and EEI thought leadership (EEI & GEO, 2021), we see a need to offer *evaluation ecosystem* as a conceptual framework and analytic tool to connect macro-level theoretical discourse around in/equity to empirical research and practical guidance that has a predominantly inter-/intrapersonal, micro-level focus. We also ground our conceptualization and use of evaluation ecosystem in firsthand accounts from practicing evaluators, as a necessary first step from which future work can build.

3.1. Ecosystem as a theoretical framework for research on evaluation

Finding no theoretical framework to understand evaluation as an ecosystem in the existing literature, we drew on several sources to initially develop our framework, which we subsequently refined iteratively and inductively as we made sense of interviewee responses. This section briefly describes key literature we drew from and synthesized.

Bronfenbrenner (1994, 1997) first developed a socioecological framework for human development in the 1970 s and 1980 s. The model illustrates how layered systems interact to influence each other and each individual's development over their lifespan. Individual and inter-/intrapersonal dynamics constitute the micro level at the center of the nested layers. Additional layers extend outward identifying broader influences (e.g., organization, community, institutions, sociopolitical and historical contexts).

We also drew on several applications of a socioecological model and recent uses of the phrase "evaluation ecosystem" within evaluation literature. Al Hudib and Cousins (2020) use an ecological framework to study evaluation capacity building in which the organization is the unit of analysis. Marra (2011) considers structural policies as complex systems and develops an evaluative framework that delineates micro, meso, and macro dimensions of change with regards to socioeconomic conditions. Beyond these uses in evaluation academic literature, the phrase "evaluation ecosystem" makes recent appearances in several AEA blog posts (Gordillo, Neuner, Josephson, & Beriont, 2020; Nolan, n.d.), specifically in relation to equitable evaluation (EEI & GEO, 2021), with the term being used to refer to foundations, grantees, community members, policymakers, evaluation consultants, and others who interdependently shape evaluation practices intended to benefit marginalized communities and improve social outcomes.

4. Evaluator interviews within a mixed methods study context

We conducted the interviews discussed in this article as part of a larger mixed methods study of equity-oriented practices among evaluators living in New England, with grant funding from the Barr Foundation. The study addressed five research questions exploring the following: (1) characteristics of evaluators (e.g., demographics, experience) and evaluation providers (e.g., firm size, mission) within the region; (2) what equity means to evaluators within their practice contexts; (3) extent to which evaluators carry out equity-related practices at different phases of an evaluation; (4) supports and barriers to equity-oriented evaluation; and (5) ways to strengthen capacity for equity-oriented evaluation. The first three questions were answered using a questionnaire and the latter three using individual interviews. We conducted an integrated analysis to answer the third research question by drawing on both questionnaire and interview data. This article focuses on the fourth and fifth research questions and primarily draws on interview data, although questionnaire data is used to characterize the

participants. Detailed descriptions of the full mixed methods study and results are publicly available in unpublished reports (Gates et al., 2021) and forthcoming manuscripts.

4.1. Research questions

In this paper, we present results addressing two questions from the larger study:

- What helps and hinders evaluators when working to center equity in their work?
- What opportunities and needs are there to build capacity of evaluators and within evaluation practice to center equity?

4.2. Participant selection

We selected interviewees intending to develop theoretical or analytical inferences about factors that influence equity-oriented evaluation. Following the logic of theoretical generalization in qualitative research (Maxwell, 2013; Eisenhart, 2009), we sought to understand key factors salient across evaluators working in different types of organizations and areas of practice. This allowed us to establish a working set of factors to be extended, refined, or refuted through future studies in other contexts (Eisenhart, 2009).

We followed Robinson's (2014) four-point framework to address the theoretical and practical concerns of sampling in interview-based research. Our "sample universe" was established through a combination of the larger mixed methods study's inclusion criteria (i.e., any professional evaluator practicing in the New England region) and the evaluators who participated (n = 81). On a questionnaire administered to AEA and AEA regional affiliate members in New England, we asked respondents whether we could contact them to participate in an interview, offering a $25 gift card incentive; the upper limit of our available sample size was determined by those who agreed (n = 33). We added maximum variation sampling (Patton, 2015) to our convenience sampling strategy in subsequent rounds of interviewing by purposively inviting participants who varied from those already interviewed with respect to demographic characteristics, state of residence, area of practice, evaluator role, and organizational type.

We conducted 21 semi-structured, individual interviews (Patton, 2015) from January 12th to February 12th, 2021. While interviewees were all geographically based in New England, they conducted evaluations in local, state, national, and international contexts. Most identify as white (17; 81%), female (15; 71%), and having senior-level experience (14; 67%). They represent external (10; 48%) and internal (7; 33%) roles, and various practice areas; about half (10; 48%) work in large and medium-sized organizations and half (11; 52%) in small/independent organizations. Their organization types are primarily community/non-profit (12; 57%) and consulting/private sector (10; 48%). Table 1 shows interviewee characteristics. Interviewees, on average, were much like our overall study sample in terms of how often they carry out equity-oriented practices at different stages of the evaluation process, with responses between sometimes and often (see Table 2).

4.3. Data generation

We drafted and finalized interview questions using an expert review (third author) and a pilot process with two evaluators not in the study. The protocol consisted of four sections: interviewee evaluation background, what equity means within the interviewee's work context, the extent to which they center equity in evaluation, and their suggestions to strengthen capacity for equity-oriented evaluation. Before the interview, we shared the protocol with interviewees and encouraged them to reflect on their responses in advance. We conducted each 50–60-minute interview via Zoom video conferencing. For each discussion, one researcher led the interview, another took notes. The interviewer periodically restated the interviewee's responses as a form of member checking (Maxwell, 2013; Creswell & Poth, 2018). We audio recorded

Table 1
Interview participant demographic characteristics and evaluation practices.

Item	N	(%)	Item	N	(%)
Race/Ethnicity	21	(100%)	Area of Practice	21	(100%)
White	17	(81%)	Social Services	9	(43%)
Asian	2	(9%)	Pre-K-12 Education	6	(29%)
Black/African American	2	(10%)	Higher Education	5	(24%)
Sex	21	(100%)	Public Health and/or Health Services	8	(38%)
Female	15	(71%)	International Development and Aid	5	(24%)
Male	4	(19%)	Organizational Development and Learning	3	(14%)
Prefer not to answer/no answer	1	(5%)	Public Policy and Analysis	2	(10%)
Intersex	1	(5%)	Criminal Justice	3	(14%)
State	21	(100%)	Informal education (e.g., out-of-school, museums)	4	(19%)
Massachusetts	11	(52%)	Workforce development	6	(29%)
Maine	3	(14%)	Something else	5	(24%)
New Hampshire	3	(14%)	Organization Size	21	(100%)
Vermont	2	(10%)	Large	6	(29%)
Connecticut	1	(5%)	Mid-sized	4	(19%)
Rhode Island	1	(5%)	Small/Independent Evaluator	11	(52%)
LGBQ Identification	21	(100%)			
No	16	(76%)	Organization Type*		
Yes	5	(24%)	Community Non- profit	12	(57%)
Evaluator Role	21	(100%)	Consulting/Private Sector	10	(48%)
External	10	(48%)	Cross-sector	4	(19%)
Internal	7	(33%)	Academia/Higher Education	3	(14%)
A mix of both	3	(15%)	Government (e.g., health department)	3	(14%)
Manager of evaluations	1	(4%)	Foundation or Funder	3	(14%)
Position	21	(100%)	Pre-K-12 Education	3	(14%)
Senior-level evaluator	14	(67%)	Environmental and Social Development	1	(5%)
Mid-level evaluator	3	(14%)	Housing and Homelessness	1	(5%)
Junior-level evaluator	0	(0%)			
Something else	4	(19%)			

Note: *Indicates a select-all-that-apply question, so % is out of 21 for each response option.

Table 2
Interviewee frequency of equity-related evaluation practices by phase.

Phase	Interviewees Mean (SD) (n = 21)	Questionnaire Respondents Mean (SD) (n = 81)
Team	2.66 (0.65)	2.48 (0.60)
Throughout	2.81 (0.65)	2.73 (0.61)
Funding & contracting	2.32 (0.61)	2.22 (0.56)
Setting questions & criteria	2.54 (0.64)	2.61 (0.63)
Data collection	2.55 (1.00)	2.64 (0.55)
Data analysis	2.66 (0.74)	2.67 (0.42)
Reporting & dissemination	2.50 (0.72)	2.88 (0.18)

Notes: This table is ordered following phases of an evaluation. Each phase included between 2 and 8items with response options from (4) almost always, (3) often, (2) sometimes, and (1) never or rarely.

and transcribed interviews using an online A.I. transcription service (Sonix.ai) followed by manual accuracy checks and corrections. We annotated transcripts for more accessible review (e.g., bolding questions, flagging sections) and shared these with interviewees, who could review, redact, and add to their interview transcripts.

4.4. Interviewees' equity definitions

Within the interviews, we asked each interviewee what equity means within their work, repeating their individual definition of equity from an open-ended question on the questionnaire and asking them to expand on that definition. In an initial analysis available in a publicly available report (Gates et al., 2021), we identified three major ways these evaluators defined equity:

- **Equity *within the evaluation process***, which includes maximizing team diversity and minimizing bias; meaningful participation of intended beneficiaries; and constructing multiculturally, contextually valid knowledge.
- **Equity as *evaluative criteria to assess intervention design and quality***, specifically in terms of reach and accessibility, differential experiences, differential outcomes, and root causes of inequity.
- **Equity as *the intended use of the evaluation process***, by, for example, challenging and expanding power dynamics and decision making, providing professional training for intervention staff, and/or building evidence about equitable interventions and social justice.

Here, we do not examine the relationship between individual definitions of equity and the helpful and hindering factors that evaluators identified.

4.5. Analytic process and framework

We analyzed interview data in two phases. To start, we closely read and discussed transcripts, developing a codebook of deductive codes based on research and interview questions and inductive codes to describe interviewee responses (Charmaz, 2014; Maxwell, 2013). Two researchers used the codebook to code several transcripts in Dedoose (www.dedoose.com), cloud-based software for analyzing qualitative and mixed methods data. As a team, we reviewed coded excerpts and clarified and resolved discrepancies. The lead researcher then combined related coded data to develop themes corresponding to each research question. These initial themes address factors identified by interviewees as helping or hindering equity-oriented evaluation practice, as well as ways to strengthen capacity for equity-oriented evaluation (reported in Gates et al., 2021). Drawing from the behavioral sciences, we use the term "factors" to refer to "forces that facilitate or impede individual, collective, or environmental change based on their level of availability" (Gilmore, 2013, para. 1).

In phase two, we conducted iterative thematic analysis using the socioecological framework we developed, continuously adapting it as our understanding of equity-oriented evaluation ecosystems evolved (see Fig. 1). The lead researcher used this framework to reorganize phase-one themes into nine domains (e.g., evaluation marketplace, evaluation workplace, evaluators) situated within three levels (macro, meso, and micro) – a process that helped refine the domains of the analytic framework, as well as our understanding and articulation of the themes (see Tables 3 and 4). The lead researcher documented theme iterations using an audit trail (Maxwell, 2013; Nowell, Norris, White, & Moules, 2017) and two other researchers reviewed and helped develop themes.

5. Results

Here we seek to situate in practice the twenty-three factors interviewees identified as supporting or constraining evaluators' potential to center equity in the evaluation ecosystems of which they are part. In keeping with the way these factors are presented in Table 4, and for ease of others' adaptation, factor-level explorations in this section are organized by level and domain.

5.1. The macro-level sociopolitical context and evaluation marketplace domains

5.1.1. Public awareness and pressure for racial justice and equity

Our interviews occurred early in 2021, amidst a global pandemic that was exposing and exacerbating numerous inequities, as well as nationwide conversations about systemic racism and racial injustice in

Table 3
Domains of socioecological framework adapted for evaluation ecosystems.

Domain	Description
MACRO level: Marketplace and sociopolitical context	
Socio-politics	Public discourse and attitudes
Evaluation marketplace	Economic factors and exchanges influencing the demand and supply of evaluations
MESO level: Organizational, educational, & professional	
Evaluation profession	Academic and professional field with norms, guidelines, standards, and culture
Evaluator education	Formal training for evaluators through academic universities, professional development workshops, etc.
Evaluation workplace	Setting and procedures where the evaluator(s) work
Evaluation contracts	Formal scope of work and agreement between the funding agency and evaluation workplace
MICRO level: Intra- and interpersonal	
Evaluation stakeholders	Relationships between evaluator(s) and program leaders, funders, intended beneficiaries, and communities
Evaluation teams	Group characteristics and ways of working of those primarily responsible for the evaluation
Evaluators	Individual characteristics and ways of working of those primarily responsible for the evaluation

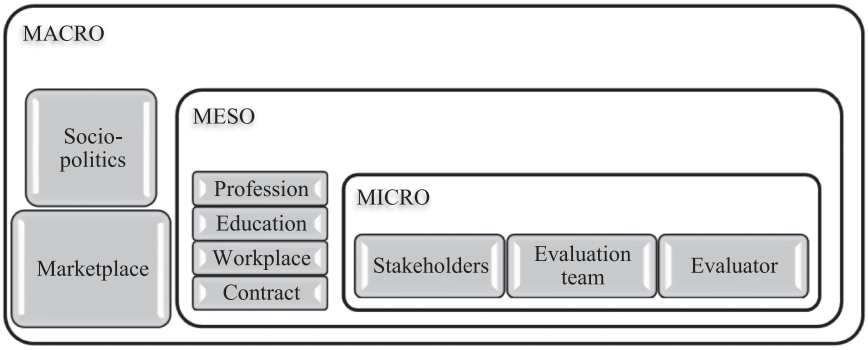


Fig. 1. Socioecological framework adapted for evaluation ecosystems.

Table 4
Factors influencing equity-oriented evaluation ecosystems.

Level	Domain	Factors
Macro	Sociopolitical Context	• Public awareness and pressure for racial justice and equity
	Evaluation Marketplace	• Market pressures and contractual tensions
Meso	Evaluation Profession	• Professional evaluation culture • Professional learning communities • Pathways for new evaluators
	Evaluator Education	• Training • Practice-based guidance and case examples • Continuous learning (not mastery)
	Evaluation Workplaces	• Values and mission alignment • Leadership and staff diversity and support • Reflexive learning culture • Confronting white racial invisibility and discomfort
	Evaluation Contracts	• Timeframe and budget • Scope of work
Micro	Evaluation Stakeholders	• Clarity and alignment around equity • Evaluators' power and agency • Engaging funders and program leaders as co-learners • Interrogate inequities in program and service designs • (Non)use of evaluation results
	Evaluation Teams	• Diversity of identities and lived experiences • Involving intended beneficiaries, communities, and new evaluators
	Evaluators	• Intersectional identity and privilege • Prioritization and conscious effort

the United States following the murder of George Floyd. Some interviewees emphasized national and global pressure that made it easier to advocate for and pursue equity-oriented evaluation. Expressing a sentiment many participants shared, one evaluator noted, "In the past year, everyone's talking about racism and inequity…. It's helped to just bring it up as a normalized piece of conversation [within their organizations and in conversations with colleagues]." Interviewees also described being personally impacted. One said, "My eyes have been opened by what's happened over the summer and the institutionalized inequities in our society." Numerous interviewees saw new initiatives and enthusiasm within their workplaces. One shared, "Our [organization's] president, after the events of the spring and the murder of George Floyd and others, launched a diversity, equity, and inclusion task force in our organization … That's been something to [show] … we're prioritizing these things." Some interviewees experienced stakeholders talking more directly about inequities and the need for change: "I think everyone's a lot more explicit across the board. I'll collect data from teens, and they talk about the inequities they observe and experience, and what they would like the organization or the school to change." Such contexts create space for evaluators to ask colleagues and program stakeholders, "Now that you know, what are you going to do about it?".

5.1.2. Market pressures and contractual tensions

Overall, while interviewees identified the sociopolitical context and public discourse around racial injustice and calls for equity as a helpful backdrop to raise questions and introduce practices around evaluation equity, they also revealed the extent to which the evaluation marketplace forces their organizations to bid for contracts and compete with other evaluation service providers for limited funding. Interviewees wrestled with where, when, and how to push for equity-oriented evaluation, weighing what might "cross the line" with funders or put them at risk of losing a contract – or their job. Interviewees described a "constant little tension of 'How much do we push? Where do we push?'"; being "very wary of losing contracts"; proposing "what we think funders are willing to go for"; and needing to "write something that we think will get funded." Some interviewees expressed frustration that pressure to secure and keep contracts outweighs commitments to advance equity within or through evaluation work. For example, though noting that their organization prioritizes working with funders and partners that share or welcome equity commitments, one interviewee shared that, "when it comes down to it," the organization remains unlikely to turn down work from "certain organizations." Even those who consistently described themselves as equity advocates found themselves making compromises based on self-preservation, "because if you say no too much, you have to think about your job." Many evaluators, particularly those in newer and smaller organizations, find that this pressure constrains and contradicts equity-oriented work.

6. The meso-level evaluation profession domain

6.1. Professional evaluation culture

Interviewees shared how evaluation norms, standards, and practices need to shift. Echoing others, one interviewee said, "we have to do some changes in our cultures of evaluation, of defining what a good evaluation is in ways that place more value on the diversity of voices and challenging perspectives that we bring in." Another interviewee specifically critiqued the predominance of "old white men who are establishing these [evaluation] frameworks," noting that while they seem "very pleased with themselves," significantly less attention is given to less familiar, more intersectionally diverse evaluators. Interviewees spoke about the influences of white supremacy and settler colonialism on evaluation; one remarked:

I think white supremacy culture is a big threat to good evaluation, because [it] tries to separate everything and act as if the only things that are valuable are things that can be measured in a certain way. And I personally have found in the community almost the exact opposite: that people, stakeholders want to understand the richer things. They don't want to be led by a tool. They want the work to lead to understanding of what they're doing.

Interviewees raised questions about specific concepts conventionally used to characterize quality evaluation, including neutrality, objectivity, validity, and expertise. One interviewee spoke about the need to interrogate the "mythology of the evaluator as somehow being this neutral vessel that takes this pristine, scientific, hermetically sealed methodology that we apply to a situation and then have clear and indisputable results." Several interviewees talked about the need to recognize and question validity as rooted in "racist, heterosexist, patriarchal" histories; others about the need to reframe validity to ensure that Black, Latinx, Indigenous, and other marginalized communities benefit from rigorous research. Relatedly, interviewees pushed back against "incredible deference to evidence-based practices" and valuing quantitative over qualitative data. Interviewees also questioned the basis for evaluator expertise, mentioning "good old-fashioned hubris," "knowing what's best from my academic training," "egos," and the "academy feeling of evaluation" as hindrances. In their reflections on expertise, one interviewee suggested that evaluators should see themselves "more like a guide or a channel for information rather than … the proclaimers of information."

6.2. Professional learning communities

Other hindrances interviewees identified within the profession are individual evaluators' isolation and competition between evaluation firms that can preclude genuine learning exchanges. Interviewees talked about how it gets "really, really lonely," especially for those who enter the field as "accidental evaluator" – a term used by some interviewees to refer to their own experience learning evaluation on the job, with no formal or intentional process for becoming an evaluator ("falling into evaluation"). Some become the "lone evaluator in an organization; an island of one." One interviewee lamented a tendency to reinvent the wheel due to "lack of opportunities to share ideas." Interviewees described seeking opportunities to learn and foster community among

evaluators and across institutions. One said, "Evaluators do not have any kind of professional learning community" and emphasized the importance of this reality. Another discussed fostering a professional network to help build evaluators' capacity and agency to do equity-oriented work:

Evaluators need more power. We talked about issues with professionalization, but if we all agree to a set of standards and ethics in which we don't fall prey to those power imbalances that pressures us to do work that doesn't address inequity, that is underfunded, along timelines that are inadequate to do our work well, evaluators in general would have more power. And if we networked effectively and provided support for those standards on a group level, we could approach evaluation negotiations with more confidence.

6.3. Pathways for new evaluators

Interviewees want the profession to mitigate "gatekeeping" and create opportunities for new evaluators. One interviewee suggested, "evaluation has to examine itself around the ways it dominates as a field and learn more from the younger up-and-coming people who are diverse in terms of race, nationality, the hemisphere of origin." Another asked, "Who is an evaluator? … I think there's a lot of gatekeeping about who is – who has the credentials or experience or whatever to undertake the traditional and nontraditional aspects of evaluation." Interviewees specifically want to develop new evaluators. One offered, "I genuinely want to know how I can help other evaluators or aspiring evaluators … from diverse backgrounds who maybe don't have the same access to the institutional supports and knowledge to advance in the field." Some saw efforts to professionalize evaluation as a potential hindrance. One confided:

I am reticent around the talks to professionalize the field and offer certificate courses and things like that, because, of course, that brings money and access to the fore. And I work with a lot of people who wouldn't have access to that, and so I worry about certification for that reason.

7. The meso-level evaluator education domain

7.1. Training

Many interviewees reflected on gaps and opportunities to better prepare evaluators for equity-oriented work within evaluator training and education in academic programs and professional development. One evaluator described being "in quant land most of the time" and wanting "alternative methods." Another wanted to learn specific approaches, methods, and processes to center equity in their repertoire, alongside other methods. Interviewees named methodological barriers they encounter frequently and their desire to learn how to work around those barriers. These include convenience samples that lack representativeness for subgroups. Small sample sizes pose risks for identifying individuals and often lack sufficient data on other characteristics to examine disparities, including LGBTQ+ identification or homelessness among youth.

Interviewees also acknowledged that learning extends beyond methods: "I don't think it has to be around the technical know-how, or how evaluation is done, but what are some ways to think about how to move through the process?" In the words of one evaluator:

Something new for me is really about why and how quantitative methods have been rooted in an understanding of statistics that has served white supremacist purposes and what we can do to use [statistics] more equitably or what should we use instead.

7.2. Practice-based guidance and case examples

Interviewees emphasized needing to learn from examples and other evaluators' efforts: "You can try to do it as much as you can. But what does it look like in action, and what does a good one look like? And what does a bad one look like? And lessons learned from that?".

Interviewees wanted "concrete steps," a sense of what "it really looks like in action," and "something more tangible." One existing resource several interviewees mentioned is teaching cases where "folks who are the implementers of evaluation or ran the programs that were the evaluand or whatever are involved in the discussion." They want narratives that cover what was done and go beyond this to reflect on what it felt like and how evaluators navigated "some of those trickier circumstances."

7.3. Continuous learning (not mastery)

Interviewees underscored how equity requires constant learning and can't be "mastered." Evaluators foresaw a "huge learning curve," particularly for white American evaluators. As an interviewee said, "it's not enough to say I recognize my privilege, but how do I work in really fundamentally different ways?" Another said:

[I]t's not something that you can finish, right? I don't want any evaluator to go into this and be like, OK, we're going to work on equity for a year, and then we're done. Check. That's just – it's not a thing. That's not how this works. It's a process. It's a mindset. And it requires a lot of humility.

Moreover, equity requires "anchoring yourself back to the field, because working out in all these multiple sectors, and it's such an interdisciplinary endeavor daily. So, grounding oneself to evaluation and then advocating with funders for better ways of doing things." Several interviewees suggested that equity be "taken seriously as one of the core competencies" for evaluators while also stating it "feels oversimplified" to believe "it's something that you can learn and be competent in."

8. The meso-level evaluation workplace domain

8.1. Values and mission alignment

Interviewees talked about being supported in carrying out aspects of equity-oriented evaluation when their organization's values and mission center equity explicitly and meaningfully. As one interviewee shared, "I've been lucky to work at places where the value is there, so they value evaluation. They value equity … so, I think having that has really helped." Several interviewees talked about how the mission and values of their organization changed over time, such as from "diversity and cultural competency" to "diversity and cultural humility" and then to "diversity, equity, and inclusion." A shift in language was viewed as a good thing when it involved authentic discussions and expression of the values driving their work. Several interviewees participated in efforts to put equity into their organization's formal commitments and saw those outcomes as starting points from which they can draw in their evaluation work.

8.2. Leadership and staff diversity and support

Interviewees underscored the importance of leadership and staff support for equity while mentioning the lack of racial and ethnic diversity as a hindrance. As one interviewee enthusiastically said, "I know I've got the backing of everyone, all the leadership. Everyone up to the CEO is going to be like 'yeah, whatever you need, do it.' And they're not going to let us make excuses not to center [equity]. I think knowing I've got that backing and support is huge." Some leaders show their support by allotting funded staff time for training, reflection, discussion, and review of practices: "The fact that we have turned it into a committee means that it's officially funded as staff time, which is helpful." While this support matters, it is insufficient if it does not include changes to hiring practices and support for leaders and staff of color. Some interviewees referenced working in "white-led" organizations, their

workplace being "whiter at the top than … at the bottom," and "people in positions of power [being] more inherently conservative" as hindering factors.

8.3. Reflexive learning culture

Interviewees shared the importance of critically reflecting on choices and decisions within the evaluation process, noting that teams can work together to promote reflexive learning. As one interviewee put it: "Evaluators need to be able to take critical perspectives on, acknowledge, and accept their own biases, and build into their evaluation work structures that address such biases." Other illustrative phrases included: "do a bunch of self-assessment," "recognize our limitations and blind spots," "honest feedback at all points in the process," "build in a reflective aspect," "have ongoing conversations about equity," "stating biases … and pulling together strategies to counter biases," and the need for "checks and balances."

8.4. Confronting white racial invisibility and discomfort

Evaluators of color and white evaluators alike talked about white colleagues' discomfort and resistance to talking about race, and a presumed predominantly white population within the New England region. While several interviewees pointed out that New England and particularly the regional and community contexts in which they were working have undergone considerable immigration and demographic shifts, one white interviewee stated, "there's a lot of confusion and a lack of understanding, in my experience, about how equity exists or doesn't exist in New England. I run a lot into claims of the predominantly white state." Interviewees voiced frustration with colleagues' rendering whiteness invisible and normative, and shared examples of palatable discomfort and resistance when trying to raise questions of diversity and equity. As one interviewee of color said:

I have questions around how do you do that [center equity] when you're working with non-diverse populations? I'm working with Caucasians. All the Caucasians work with other Caucasians. They're not even thinking about anyone that doesn't look like them, so how can I talk about equity and inclusion and diversity within these contexts?

A white interviewee shared that white colleagues assume that "we don't have to look at race because [program participants] are all youth of color." These quotations illustrate presumptions of racial homogeneity as well as a lack of attention to intersectionality within racialized groups. Interviewees spoke about wanting white colleagues to be more aware and comfortable talking about race and inequity and be willing to interrogate their own assumptions as well as assumptions within the intervention design and context through (and throughout) the evaluation process.

8.5. The meso-level evaluation contracts domain

Formal agreements between evaluation firms and those commissioning evaluations (e.g., requests for proposals, proposals, funded contracts) are critical to whether evaluators have the parameters to center equity.

8.6. Timeframe and budget

Evaluators reported "ridiculously tight time frames" and "very tight budgets," as factors that hinder centering equity in evaluations. Such conditions, according to one interviewee, result in their "constantly hav [ing] to narrow who gets to speak." One interviewee working internationally memorably described how timeframes and budget constraints can impact an evaluation:

Like with all evaluative work, no one remembers to tender it soon enough. And you have to evaluate, you know, three-million-dollar programs in four and a half weeks. And we only have a budget to send you to the field for six days. I'll do what I can, but there's a limited sort of scope there to come up with anything good or useful … it becomes … very much a box-ticking exercise.

One interviewee concluded, "I think equity matters, and we're always trying to make it better in a context of deadlines (and) budgets."

8.7. Scope of work

Throughout the interviews, evaluators working on regional, national, and international projects large and small discussed how scope of work often constrained their potential to center equity in evaluation designs and processes. Concerns included the extent to which scopes were "scripted," limiting evaluators' autonomy and capacity to shape evaluation questions and methods. As one interviewee put it, scopes of work bind evaluators to "evaluate in the way [funders] want you to evaluate and measure what they say should be measured." Naming incarceration rate, graduation rate, and unplanned pregnancy as problematic examples of "what some funders [of youth services] want to look for," one internal evaluator noted:

Those are big markers, but our young people are in experiences where some of those might be out of their hands. If that's going to be a value of the program's success, that's so limiting. It makes people numbers. It doesn't look at other things in a person's life, and it doesn't look at their communities. It's looking at the wrong stuff.

While several interviewees recalled times they tried to negotiate suggested scopes of work in some way, with few exceptions their own proposed scopes got "scaled back, rolled back, pushed back, or canceled."

9. The micro-level evaluation stakeholders domain

9.1. Clarity and alignment around equity

Interviewees shared that differences, and assumptions of sameness, regarding what equity means are a hindrance within and among their teams, workplaces, and clients. As one interviewee said, "Well, I know how I conceive of equity, but if I ask that of every client I have now, I'd probably get five different definitions." Another asked, "how do we learn to hold our different ways of seeing and thinking [about equity] contingent?" The lack of a shared definition of equity poses various challenges. Several interviewees discussed facing pushback around expanding client or program definitions to include structural or systemic reform. One interviewee reflected on how understanding equity as access to resources and opportunities versus sustaining cultural worldviews would have dramatically different implications for an evaluation of a science program in an Indigenous community. Another, who spoke about the "rights of nature" as an evaluation-related equity issue, was the only of the twenty-one interviewees to include the environment in conceptions of equity.

9.2. Evaluators' individual sense of power and agency

Interviewees' beliefs differed regarding how much power and agency they feel they have to center equity within evaluations. While one interviewee stated, "I think evaluators feel like they don't have a lot of power to design their work because of the funders," others noted that experience and seniority within their organization had helped them learn to negotiate with funders. Several interviewees suggested that lead evaluators have greater power and agency, whereas newer evaluators or those in support roles (e.g., research assistants, data analysts) may "feel like it's all out of their control." This can be a frustrating experience: "I end up feeling extremely powerless, a lot. I'm a relatively new evaluator, feeling like I can't advocate for all the things that I want to."

9.3. Engaging funders and program leaders as co-learners

Interviewees spoke about the hierarchical and limited communication between funders and evaluators as a hindrance. One interviewee said:

This work and even the permission to do this work is so contingent on a relationship with the funders and relationship with the programs. 100% focusing on that relationship very early in the process is the only way I can see this ever being constructive.

Evaluators want open dialogue with funders. They contended that equity requires funders to shift their thinking about what "success" and "programs working" mean to better address "the multi-pronged causes and effects of, say, poverty or poor health." Such foci require longer-term, bigger-picture thinking about program goals and design, as well as evaluation. One interviewee described funders relying on the evaluation team to "teach them" about equity, noting that such dynamics and expectations fall far short of "a more married and symbiotic [funder–evaluator] relationship."

9.4. Interrogating inequities in program and service designs

Many interviewees discussed unchallenged assumptions and inequities within the conceptualization, design, planning, and implementation of evaluands, and how to raise these issues in productive ways. As one interviewee put it, "we're only starting to grapple with – so, there is inequity. What do we do?" Interviewees called attention to a recurring experience of working with funders and program leaders who "already have in their head the solution." This narrows funders' and leaders' views of intended beneficiaries' and communities' assets and needs, which limits what they ask for and want to learn in the evaluation process. Interviewees know that raising critical questions about root causes of inequity and interrogating programmatic assumptions are essential steps but described being unsure how to do so:

I think the difficult thing for us as evaluators is pushing back and challenging and figuring out how to do that … when we see inequity that's built into programming and figuring out how to raise attention and awareness to that.

Interviewees were also sensitive to how political and economic contexts force them to frame problems and interventions to demonstrate the appearance of progress within short timeframes, norms that disincentivize longitudinal and formative evaluation approaches.

9.5. (Non)Use of evaluation results

Interviewees were frustrated and discouraged by stakeholders who ignore evaluation results that challenge their assumptions about inequities. Many depicted futile realizations that specific clients and intervention leaders denied, downplayed, or ignored evaluation results – responses that perpetuated power imbalances and continued reproducing inequities. One interviewee said, "People think, oh, the evaluation will just show what I already know to be true, and that very rarely happens in its totality. And that's where the shock and denial sometimes come in."

Another stated: "They just don't care. We could get some information that would be interesting and actionable around the ethnic and racial composition of the people they're serving, and they're just not doing anything with it." These quotes point to the importance of intervention leaders being willing to question their own assumptions, beliefs, and visions. When this doesn't happen, evaluators question the potential of evaluation to address inequities.

10. The micro-level evaluation teams domain

10.1. Diversity of identities and lived experiences

Interviewees emphasized evaluation team composition as a factor influencing equity-oriented evaluation. One interviewee enthusiastically shared:

I'm super excited that we have a team that is more than half people of color. It also includes a co-PI who has a disability. It has multiple team members who identify as LGBTQ. We've been able to diversify our group through partnerships.

Others mentioned limited diversity on their teams as a hindrance. Some were strategizing within their organizations to broaden their networks and form new partnerships to expand perspectives within their teams.

10.2. Involving intended beneficiaries, communities, and new evaluators

Evaluators working internationally and with beneficiary communities distinct from their teams discussed the importance of involving people familiar with the linguistic, cultural, and geopolitical context. Using phrases like "more local voices at every stage," "hiring an evaluator that brings a dimension that I don't have," "bring on an adjunct team of folks who are not professional evaluators," "tapping into talent that we're not already aware of," and "encourage new people to add their voice," interviewees reflected on rethinking their own leadership and power within the team (e.g., prioritizing not being the lead), how to bring community members and locals into evaluation processes, and how to go beyond the evaluation to build capacity and offer research and evaluation training.

11. The micro-level evaluators domain

11.1. Intersectional identity and privilege

Interviewees spoke about how their own intersectional identities and experiences of both discrimination and privilege influence how they consider equity in practice. Several evaluators of color experienced tokenization and expectations that they will lead equity work. For example, one evaluator of color spoke about being brought onto a project "because the project deals with race stuff" and others "don't have caché to enter those communities." Whereas some evaluators of color described facing comparatively higher expectations for doing equity work and/while being undervalued for their roles, some white evaluators wrestled with privilege and uncertainties about their level of responsibility:

What is my role as a white woman in this space for promoting equity? And how much good do I do in the role that I'm in, and how much good would I do if I stepped aside and gave space to somebody else?

Participants predominantly focused on race in relation to discrimination and privilege, but also mentioned other aspects of identity including gender, sexual orientation, ability, culture, and language.

11.2. Prioritization and conscious effort

Many interviewees discussed how promoting equity takes consistent prioritization and conscious effort. One interviewee reflected:

The biggest and most humbling thing is just how much we need to keep it front of mind and conscious. Because if you're not consciously paying attention, it's very easy to slip and miss things…. Just because you're intellectually aware that this is important, that doesn't mean that at the moment when you're busy, you're paying attention.

Several interviewees shared recent stories about falling short of their aspirations; one confided, "When I'm stressed out, it's very easy to just go on autopilot and say, OK, I need to crank out this data collection. I need to crank out the survey draft." Interviewees emphasized the need to slow down and allow time and space to question conventional methods and try alternatives.

12. Discussion

Our study extends prior calls for changes by individual practitioners and within the evaluation profession to center equity. Results highlight multiple factors and dynamic conditions that shape how, by whom, and with what consequences evaluations are (and can be) designed and conducted to center equity.

12.1. Macro level

At the macro level, our results underscore what many presumed – public attention on racial justice and equity created some space for shifts in evaluation work. A mix of public discourse, news coverage, and organizational position and value statements all contributed to a backdrop in which some evaluators found openings to discuss equity within evaluation work. Our results also highlight the substantial influence of the evaluation marketplace, in keeping with research discussing the growing domination of the federal evaluation market by large external firms and multi-year evaluations (Lemire, Fierro, Kinarsky, Fujita-Conrads, & Christie, 2018b), and work arguing that small(er) evaluation providers face unique challenges in this landscape (Hwalek & Straub, 2018). These market forces create ongoing and intensifying competitive pressure to secure and keep contracts, influencing how and by whom evaluations get carried out.

12.2. Meso level

At the meso level, many evaluators practice with little or no formal training in equity-oriented approaches, within research and evaluation cultures often at odds with equity principles. Moreover, competition between evaluation workplaces and the resultant isolation and lack of exchanges between evaluators can restrict the pace of learning. In line with prior research, our results point to needed shifts in evaluator education and pathways for new evaluators, some of which are underway, including trainings (e.g., We All Count) and initiatives to support evaluators of color and from minoritized backgrounds (e.g., AEA's Graduate Education Diversity Internship, Expanding the Bench's Leaders in Equitable Evaluation and Diversity). Evaluators emphasized the value of frameworks, particularly the Equitable Evaluation Framework™ (Dean-Coffey, 2017) and practice guides (e.g., Andrews et al., 2019). However, they also want to read case examples and hear what evaluators are trying and learning. One pathway to address these needs that prior research has overlooked and our results highlight is professional learning networks. As interviewees expressed, evaluators want to talk with each other about their practices in non-competitive, supportive, and growth-oriented ways.

Our results also highlight how evaluation workplaces and stakeholder relationships can shape the conditions for equity-oriented evaluation. Evaluators are not autonomous professionals; they work under supervisors and within organizations with their own cultures, standards, and review processes for evaluation. Organizational values and missions, and teams of intersectionally diverse and supportive leadership and staff, can collectively produce organizational cultures that prioritize and practice equity as integral, enacted by norms like providing time for training and equity audits of evaluation documents. It is important that evaluation contracts include adequate timeframes, budgets, and scopes of work aligned with equity (Lo & Espiritu, 2021), practices interviewees attested to rarely seeing. Additionally, stakeholders, particularly funders and program leaders, can deter or enable equity-oriented evaluation depending on their own understandings of what equity means and their own (de)valuing of evaluation practices that center equity.

Our results reveal potential to extend prior research in the context of equity with regards to the risk of conflicts of interest between those who commission and conduct evaluations (Datta, 2016; House, 2011), as well as evaluation nonuse and misuse. Results also reveal gaps in current guidance, especially regarding how evaluators navigate different conceptions of equity and draw attention to missing considerations within these conceptions (e.g., environmental justice), and how evaluators can productively question inequities in ways that stakeholders hear and act on.

12.3. Micro level

At the micro-level, prior research focuses on individual evaluators' positionalities, critical self-reflection, and cultural competencies. Our results found most evaluators work in teams, yet within evaluation scholarship little is understood about evaluation teams, their compositions and dynamics, or their collective competencies and reflexive practices. Interviewees highlighted the importance of intersectional diversity and different lived experiences within the teams they are part of, which they noted include community members and new evaluators, all of which corresponds to factors shared in prior work (AEA, 2011). Our results highlight how teams can support a "high degree of self-awareness and self-examination" (AEA, 2011) or reflexivity (EEI & GEO, 2021) and facilitate continuous learning or "plasticity" (i.e., the ability to change in response to new experience and ideas) (EEI & GEO, 2021). Interview data revealed how organizations and teams can shape the time and norms for self-scrutiny and for addressing cultural and other biases in safe and productive ways. This is particularly important for those who had experienced or expressed concern about their disproportionate risk when drawing awareness to biases and injustices. Our results also identify the importance of evaluator willingness to challenge white racial discomfort. This connects with House's (2017) work on the white racial frame and culturally responsive evaluation generally (Hood, Hopson, & Kirkhart, 2015), and suggests a need to further examine how white evaluators and evaluators of color may address pushback from colleagues differently.

Finally, at the individual level, our work supports prior scholarship and efforts that call for racially and ethnically diverse evaluators within the field (Lo & Espiritu, 2021), while also mitigating inequities for these evaluators within evaluation and organizational contexts. Results underscore the assets evaluators of color bring to evaluation through their personal histories, lived experiences, and levels of awareness (Lo & Espiritu, 2021). However, they also raise concerns about the ways in which the contributions of evaluators of color may get devalued or appropriated. Examples discussed include expecting evaluators to show up at program events to connect with racially minoritized participants – an ask outside the evaluation scope and budget.

12.4. Implications

This final section discusses contributions, limitations, and future research directions.

12.5. Contributions

Equity is emerging as a normative ideal for the evaluation field and professional responsibility for evaluators. However, we are in an unprecedented, precarious moment of reckoning with centuries of social and ecological injustice and inequity – a watershed moment in which "equity" is at risk of being subsumed in dominant neoliberal accountability culture and turned into an empty signifier or, as a few interviewees noted, another "box to check." This study moves beyond prior theoretical literature that locates change at the level of individual practitioners and the profession, providing an empirical basis for arguing that change must extend across evaluation ecosystem domains. The study contributes one of the first conceptual understandings of evaluation as an ecosystem and does so by drawing empirically on practitioner perspectives. The framework delineates important units of analysis/domains for change. The study also identifies equity-influencing factors of potential relevance to other geographical

regions and cultural contexts. To support readers in exploring the transferability of our findings to their contexts, we recast identified factors as a set of questions (see Table 5) that can be used by evaluators, program leaders, and others to consider factors that might be leverage points for change within their contexts, and can complement those that focus on the project phase (Cerna et al., 2021).

Table 5
Questions examining conditions for equity-oriented evaluation ecosystems.

Level	Domain	Factors
Macro	Sociopolitical Context	• What societal sources of inequity are being reproduced in program and evaluation contexts? What societal pressures to address in/equity are present in these contexts?
	Evaluation Marketplace	• How do supply of and demand for evaluations, and competition between evaluation providers help or hinder equity?
Meso	Evaluation Profession	• What, if any, aspects of professional evaluation culture conflict with equity? How can dominant evaluation constructs be interrogated, problematized, and reframed? • Are evaluators (dis)incentivized to share practices and learn together across organizations and sectors? What learning networks (could/should) exist? • Are there clear pathways to support new evaluators? How can organizations and teams contribute to these pathways?
	Evaluator Education	• In what trainings have/could/should evaluators participate(d)? • How can emerging guidance and case examples be used to inform discussions and decisions about evaluation practice? • Is a culture of learning or of procedural "mastery" present?
	Evaluation Workplaces	• Do the organizational values and mission meaningfully include commitments to equity? Are they equity-aligned or at odds? • Do organizational leaders and staff embody cultural diversity and enact their power to support equity-oriented work? • Does the organizational culture promote reflexivity and learning, including establishing dedicated time for it? • Are those with power and privilege pushed to confront and learn from the discomfort inherent in anti-oppressive work?
	Evaluation Contracts	• Are timeframes and budgets adequate for centering equity in evaluations? Do scopes of work explicitly address in/equity?
Micro	Evaluation Stakeholders	• In what way(s) is equity defined and to what extent do definitions align across stakeholders? What spaces exist for differing perspectives to be heard, respected, and negotiated? • What is the nature of relationships between funders, commissioners, program leaders, and evaluators? Is sharing power and co-learning occurring or an acceptable possibility to those with disproportionate power? • Are program leaders open to questioning assumptions, beliefs, and visions, and acting on evaluation results?
	Evaluation Teams	• Do teams have members with diverse intersectional identities and lived experiences? • Are there ways for those in the intended beneficiary community and new evaluators to participate meaningfully on teams? Are they valued and fairly compensated?
	Evaluators	• How do evaluators' identities, lived experiences, and privileges influence their work? • To what extent do evaluators exhibit critical self-awareness in their work with others? • To what extent do individual evaluators prioritize and seek opportunities to learn and grow?

12.6. Limitations

We intentionally addressed the main limitations of this study – its focus on one regional context and point in time, the exclusion of non-evaluators' perspectives, and the majority of interviewees being white and female – by trying to maximize variation in our selection of interviewees in other ways, including areas of practice, and focusing our analysis on theoretical, not statistical, generalization: "the attempt to develop a refined understanding of a generic process" (Eisenhart, 2009, p. 60). We leave readers to explore the relevance and transferability of the focal domains and factors in their unique contexts and future work. Another limitation of our work is the use of a hierarchically nested, static conception of an evaluation ecosystem (Fig. 1) which masks the lateral and dynamic interplay of factors. We welcome more nuanced future work.

12.7. Future directions

Given the limited empirical research on equity in evaluation, there are tremendous possibilities in what researchers might explore. Here we identify several directions we view as promising and draw on systems concepts of interrelationships, perspectives, and boundaries to do so (Williams, 2015; Gates, 2021).

12.8. Interrelationships and dynamics shaping in/equity in evaluation ecosystems

Future studies to examine evaluation ecosystems can adapt, test, and refine the framework proposed here and the specific domains and factors identified. This work could explore questions such as: What constitutes an evaluation ecosystem? Which actors and factors influence evaluation practice, and how so? Which and how, if at all, do domains and factors identified in this study translate to other evaluators and geographical contexts? Future work can also shift beyond identifying factors of influence at different levels of evaluation ecosystems toward theorizing and empirically examining how the interplay of various factors shape the ways and extent to which evaluations center equity and the professional experiences of evaluators. This work might ask: How do factors interrelate and generate dynamics that reinforce or challenge inequities within particular contexts?

12.9. Perspectives on in/equity and evaluation

Future research should also incorporate other viewpoints, particularly those of funders, commissioners, program leaders, and participants involved in and affected by evaluands and evaluations. Questions to explore include: How do different stakeholders frame 'equity,' 'evaluation,' and the relationship between the two? Where do different framings locate the root causes of inequity and levers for change? Where and how do perspectives differ or conflict?

12.10. Normative bases for in/equity in evaluation

Finally, future research should move beyond descriptive and explanatory understandings of equity within evaluation ecosystems to engage normatively with what "quality" evaluation means, critically surface and interrogate boundary judgments (Williams, 2015; Gates, 2018), and raise questions such as: Which worldviews are (and should be) valued and marginalized with regard to equity in evaluation? To what extent is there (should there be) space for different moral and political arguments for equity within evaluations? What are (and should be) the professional responsibilities of evaluators regarding equity? What are the normative limitations of 'equity' as compared to justice, liberation, self-determination or other moral-political visions?

13. Conclusion

Nationally and globally, there are concerns and promising developments regarding evaluation and evaluating that seek to challenge inequity and promote equity. This study identified multi-layered factors that influence how and the extent to which evaluators can center equity within evaluation practice. The factors suggest ways different evaluation ecosystem actors can shape conditions for/of equity-oriented evaluation. As one of the first studies on equity within research on evaluation, and the first to use a socioecological conceptual framework, this study offers foundations and starting points for what we anticipate will become a robust body of research as the field continues to discern and transform itself to pursue equity.

Funding

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

CRediT authorship contribution statement

Emily F. Gates: Conceptualization, Methodology, Investigation, Writing – original draft preparation, Project administration, Supervision. **Joseph Madres**: Conceptualization, Investigation, Writing – original draft preparation. **Jori N. Hall**: Methodology, Writing – review & editing, Supervision. **Kayla Benitez Alvarez**: Investigation, Writing – review & editing.

Acknowledgments

We thank the twenty-one evaluators who graciously shared their time and candid experiences with us as interview participants; the four anonymous reviewers whose feedback strengthened the paper; our colleague, Eric Williamson, who supported the questionnaire portion of the larger study; and Tawnya Fay for her clear and thoughtful editorial work.

References

Al Hudib, H., & Cousins, J. B. (2020). Evaluation policy and organizational evaluation capacity building: Application of an ecological framework across cultural contexts. Journal of Multidisciplinary. *Evaluation, 16*(36), 37–56.

American Evaluation Association. (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. ⟨https://www.eval.org/About/Guiding-Principles⟩.

Anderson, M. A., & Mastri, A. (2021). Culturally responsive and equitable evaluation for federal evaluation staff. *Mathematica*. ⟨www.mathematica.org/publications/culturally-responsive-and-equitable-evaluation-for-federal-evaluation-staff⟩.

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*.

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

Boyce, A. S. (2017). Lessons learned using a values-engaged approach to attend to culture, diversity, and equity in a STEM program evaluation. *Evaluation and Program Planning, 64*, 33–43. https://doi.org/10.1016/j.evalprogplan.2017.05.018

Bronfenbrenner, U. (1981). *The Ecology of Human Development: Experiments by Nature and Design*. Harvard University Press,.

Bronfenbrenner, U. (1994). Ecological models of human development. In *International Encyclopedia of Education* (2nd ed..,, Vol. 3). Elsevier,.

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

Center for Evaluation Innovation. (2017). Equitable evaluation project framing paper. ⟨www.equitableeval.org/blog-main/2017/7/17/equitable-evaluation-framing-paper⟩.

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

Change Elemental. (2021). Deep equity. ⟨https://changeelemental.org/offerings/deep-equity/⟩.

Charmaz, K. (2014). *Constructing Grounded Theory* (2nd edition.,). SAGE.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches* (4th ed..,). SAGE,.

Datta, L. E. (2016). Who pays the piper: Funding sources and evaluation's contribution to equity. In S. Donaldson, & R. Picciotto (Eds.), *Evaluation for an Equitable Society* (pp. 153–168). Information Age Publishing.

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

Dean-Coffey, J. (2018). What's race got to do with it? Equity and philanthropic evaluation practice. *American Journal of Evaluation, 39*(4), 527–542.

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

Eisenhart, M. (2009). Generalization from qualitative inquiry. In K. Erickan, & W.-M. Roth (Eds.), *Generalizing from Educational Research: Beyond Qualitative and Quantitative Polarization* (pp. 51–66). Routledge.

Equitable Evaluation Initiative (EEI) & Grantmakers for Effective Organizations (GEO). (2021). Shifting the evaluation paradigm: The Equitable Evaluation Framework. ⟨www.equitableeval.org/blog-main/shifting-the-evaluation-paradigm-the-equitable-evaluation-framework-eef⟩.

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⟩.

Forestieri, M. (2020). Equity implications evaluating development aid: The Italian case. *Journal of Multidisciplinary Evaluation, 16*(34), 65–90.

Garibay, C., & Teasdale, R. M. (2019). Equity and evaluation in informal STEM education. In A. C. Fu, A. Kannan, & R. J. Shavelson (Eds.), *New directions for evaluation, 161* pp. 87–106). Wiley Periodicals, Inc.. https://doi.org/10.1002/ev.20352

Gates, E. F., Walton, M., Vidueira, P., & McNall, M. (2021). Introducing systems- and complexity-informed evaluation. In E. F. Gates, M. Walton, & P. Vidueira (Eds.), *Systems and Complexity-Informed Evaluation: Insights from Practice. New Directions for Evaluation* (pp. 13–25). https://doi.org/10.1002/ev.2046

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 for equity in New England evaluation (SCENE) short report. https://barrfdn.issuelab.org/resource/strengthening-capacity-and-equity-in-new-england-evaluation.html.

Giacomini, M., & Hurley, J. (2008). Issues in evaluating equity. *Health Promotion Evaluation Practices in the Americas* (pp. 285–298). Springer,.

Gilmore, G. D. (2013). Enabling factors. *Oxford Bibliographies*. ⟨www.oxfordbibliographies.com/view/document/obo-9780199756797/obo-9780199756797-0081.xml⟩.

Gordillo, L., Neuner, R., Josephson, L., & Beriont, L. (2020). Evaluation as an ecosystem: Building our ability to collaborate effectively. *American Evaluation Association*, 365. October 9) ⟨https://aea365.org/blog/evaluation-as-an-ecosystem-building-our-ability-to-collaborate-effectively-by-liz-gordillo-rory-neuner-leah-josephson-and-lauren-beriont/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+aea365+%28AEA365%29⟩.

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

Greene, J. C. (2016). Advancing equity: Cultivating an evaluation habit. In S. I. Donaldson, & R. Picciotto (Eds.), *Evaluation for an equitable society*. Information Age.

Hall, M. E. (2018). Evaluation's Race Problem in the United States: A Call to Action for the Profession and the American Journal of Evaluation. . *American Journal of Evaluation, 39*(4), 569–583. https://doi.org/10.1177/1098214018792624

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*. ⟨https://www.aisp.upenn.edu/centering-equity/⟩.

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. *Handbook of practical program evaluation*, 281.

Hopson, R., Kirkhart, K. E., & Bledsoe, K. L. (2012). Decolonizing evaluation in a developing world: Implications and cautions for equity-oriented evaluation. *Evaluation for equitable development results*, 59–82.

House, E. R. (2017). Evaluation and the framing of race. *American Journal of Evaluation, 38*(2), 167–189. https://doi.org/10.1177/1098214017694963

House, E. R. (2011). Conflict of interest and Campbellian validity. *New Directions for Evaluation, 130*, 69–80.

House, E., & Howe, K. R. (1999). *Values in Evaluation and Social Research*. SAGE.

Hwalek, M. A., & Straub, V. L. (2018). The small sellers of program evaluation services in the United States. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The Evaluation Marketplace: Exploring the Evaluation Industry* (pp. 125–143). Wiley Periodicals, Inc.

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

Lemire, S., Nielsen, S. B., & Christie, C. A. (2018a). 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

Lemire, S., Fierro, L. A., Kinarsky, A. R., Fujita-Conrads, E., & Christie, C. A. (2018b).). The U.S. federal evaluation marketS. B. Nielsen, S. Lemire, & C. A. Christie (Eds.). *New Directions for Evaluation, 160*, 63–80.

Letiecq, B. L., & Bailey, S. J. (2004). Evaluating from the outside: Conducting cross-cultural evaluation research on an American Indian reservation. *Evaluation Review, 28*(4), 342–357.

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*.

Lucas, K., van Wee, B., & Maat, K. (2016). A method to evaluate equitable accessibility: Combining ethical theories and accessibility-based approaches. *Transportation, 43*(3), 473–490. https://doi.org/10.1007/s11116-015-9585-2

Mathison, S. (2016). Confronting capitalism: Evaluation that fosters social equity. In S. Donaldson, & R. Picciotto (Eds.), *Evaluation for an equitable society* (pp. 83–108). Information Age Publishing.

Marra, M. (2011). Micro, meso, and macro dimensions of change: A new agenda for the evaluation of structural policies. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution, and beyond. Comparative Policy Evaluation* (p. 18). Transaction Publishers.

Maxwell, J. (2013). *Qualitative Research Design: An Interactive Approach* (3rd ed..,). SAGE.

Mendez, K., & Taniuchi, A. (2020). Expanding the Bench™ week: Culturally responsive and equitable evaluation: What is it and why is it important? *American Evaluation Association, 365*. ⟨https://aea365.org/blog/expanding-the-bench-week-culturally-responsive-and-equitable-evaluation-what-is-it-and-why-is-it-important-by-karla-mendez-alina-taniuchi/⟩.

Nolan, C. (n.d.). It's time to let go of tired narratives about talent in evaluation. Center for Evaluation Innovation. ⟨www.evaluationinnovation.org/insight/its-time-to-let-go-of-tired-narratives-about-talent-in-evaluation/⟩.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1). https://doi.org/10.1177/1609406917733847

Patton, M. Q. (2015). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice* (4th ed..,). SAGE.

Peck, L. (2018). The big evaluation enterprises in the United States. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The Evaluation Marketplace: Exploring the Evaluation Industry* (pp. 97–124). Wiley Periodicals, Inc.

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

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology, 11*(1), 25–41. https://doi.org/10.1080/14780887.2013.801543

Rogers, P. J. (2016). Understanding and supporting equity: Implications of methodological and procedural choices in equity-oriented evaluations. In S. Donaldson & R. Picciotto (Eds.). *Evaluation for an Equitable Society*. Information Age Publishing,.

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⟩.

Stake, R. E. (2016). Finding merit, diminishing hurtful discrimination, with less pursuit of meritocracy. In S. Donaldson, & R. Picciotto (Eds.), *Evaluation for an Equitable Society* (pp. 181–198). Information Age Publishing.

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

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*.

Williams, B. (2015). Prosaic or profound? The adoption of systems ideas by impact evaluation. *IDS Bulletin, 46*(1), 7–16. ⟨https://bulletin.ids.ac.uk/index.php/idsbo/article/view/120⟩.

Emily F. Gates, PhD, is an assistant professor in the Measurement, Evaluation, Statistics, and Assessment department at Boston College. She works as an evaluator, social scientist, and mixed methodologist. Her research focuses on the role of evaluation in addressing complex problems and changing systems. Driven by a democratic vision for evaluation, she advances evaluation theory, methods, and practice that use systems thinking and approaches, make values explicit, and center equity. She recently co-authored a book, *Evaluating and Valuing in Social Research* (Guilford Press) and guest co-edited a special issue of *New Directions for Evaluation* exploring systems approaches to evaluation. Gates holds a PhD and Ed.M from the University of Illinois, Urbana-Champaign.

Joseph Madres, is a PhD student in the Teaching, Curriculum, and Society department in the Lynch School of Education and Human Development at Boston College. Before returning to graduate school, Joe was the Director of Grants and Research at the Center for Creative Education in West Palm Beach, Florida, where he worked collaboratively to design, implement, evaluate, and pursue grant funding for an array of arts-integrated education programming for partner schools and the local community. Prior to his work at CCE, Joe was a middle and high school art teacher.

Jori N. Hall, PhD, is a multidisciplinary researcher and evaluator focused on social inequalities and the overall rigor of social science research. Specifically, her work addresses issues of evaluation and research methodology, cultural responsiveness, and the role of values and privilege within the fields of education and health. She is the author of *Focus Groups: Culturally Responsive Approaches for Qualitative Inquiry and Program Evaluation*. She is an associate editor for the *American Journal of Evaluation* and the co-editor of a special issue on mixed methods inquiry in the International Journal of Research and Method in Education. Dr. Hall holds a B.S. from Bradley University, a Master's Degree from DePaul University, and a Ph.D. from the University of Illinois, Urbana-Champaign.

Kayla Benitez Alvarez, is an undergraduate student and McNair scholar in the Lynch School of Education and Human Development at Boston College studying Applied Psychology & Human Development. Her research focuses on college access and equity for underrepresented students in higher education. In the future she hopes to obtain her PhD in an education related field.